UNIVERSITY OF SUSSEX

SCHOOL OF MATHEMATICS AND PHYSICAL SCIENCES DEPARTMENT OF MATHEMATICS

A Prediction Model for Identifying Adolescent Substance Use Risk Based on Mental Health and Peer Influence

Student: Candidate Number: 260801

Agalya Muthukrishnan Supervisor: Prof Konstantin Blyus

Submitted in partial fulfillment of the requirements for the MSc degree in Data Science at the University of Sussex

September 2023

Acknowledgements

I would like to express my heartfelt gratitude to my dedicated supervisor, Konstantin Blyss, for his unwavering support, invaluable guidance, and boundless patience throughout the journey of this research. His expertise and insights have been instrumental in shaping this dissertation.

I am also grateful to Dr. Darya, the esteemed psychologist whose expertise illuminated the path of my research. Her valuable input and thoughtful guidance have been pivotal in enriching the understanding of the complex interplay between adolescent alcohol consumption, mental health, and social influences.

This dissertation was inspired by my senior, Megan Jennefer Robertson, whose mentorship and shared wisdom paved the way for this research endeavor. I am deeply indebted to her for her encouragement, assistance, and inspiration.

I extend my sincere appreciation to my friends and family who stood by me with unwavering encouragement throughout this academic journey. Your belief in my abilities was a constant source of strength.

Thank you to each one of you for your roles in making this research a reality.

Content Page

1.0 Introduction

Mental health issues among adolescents have emerged as a critical public health concern on a global scale, transcending geographical, cultural, and socioeconomic boundaries. Recent statistics paint a sobering picture, revealing that approximately 20% of adolescents encounter mental health challenges within any given year [1]. What's even more disconcerting is that a significant proportion of these issues seem to take root during the formative years of early adolescence. Astonishingly, nearly 50% of mental health problems manifest by the age of 14, a statistic that balloons to a staggering 75% by the age of 24 [2]. The scope of this issue is not confined to late adolescence and early adulthood; rather, it extends its reach into even younger age brackets. Around 10% of children and young individuals aged 5 to 16 years are diagnosed with clinically significant mental health problems [3]. However, the alarming reality is that a substantial percentage of these young individuals do not receive the timely and adequate interventions required [4].

This dissertation embarks on a journey to delve into the intricate web of relationships that exist between alcohol consumption during adolescence, mental health outcomes, and the pervasive influence of one's social milieu. It represents a comprehensive exploration grounded in the rigorous analysis of data obtained from the Millennium Cohort Study, a groundbreaking longitudinal study conducted in the United Kingdom. The primary aim is to uncover valuable insights into the complex interplay between early alcohol consumption and mental health outcomes among adolescents aged 11 to 17. Through an extensive examination of diverse studies, this review seeks to identify recurring patterns, emerging trends, existing controversies, and the critical gaps that permeate the current body of literature. Ultimately, this understanding will contribute to more informed decision-making processes in the development of mental health interventions and prevention strategies designed specifically for adolescents.

1.1 Scope of the Issue

Adolescent mental health is a multifaceted phenomenon encompassing a myriad of challenges such as anxiety disorders, depression, substance abuse, and behavioral disorders. While the roots of these issues can often be traced back to childhood, adolescence is undeniably a pivotal period in the life course where vulnerabilities may escalate, and opportunities for timely intervention can be harnessed [5]. Substance abuse, particularly alcohol consumption, stands out as a significant risk factor for the emergence and exacerbation of these mental health challenges. As young individuals grapple with the transitions and transformations that come with adolescence, the allure of alcohol and its potential as a coping mechanism can become increasingly enticing.

1.2 Research Rationale

The rationale for this research endeavor is grounded in the urgent need to address the burgeoning crisis of adolescent mental health. While substantial attention has been directed towards understanding the individual contributors to mental health issues, a more holistic examination is essential. This dissertation contends that to fully comprehend the complexities of adolescent mental health, one must consider the role played by alcohol consumption as well as the influence exerted by peers and family. The overarching goal is to dissect these intricate relationships and uncover evidence-based insights that can guide interventions.

1.3 Research Questions

The central research question that guides this study is as follows:

- Does early alcohol consumption during adolescence, as well as the influence of peers and family, significantly impact mental health outcomes among adolescents aged 11 to 17?
- How do parental and peer influences impact adolescent alcohol and drug use?
- Is there a relationship between early alcohol consumption and mental health outcomes at age 11 and 14?

1.4 Dissertation Structure

The remainder of this dissertation is organized as follows: Chapter 2 presents a comprehensive review of the existing literature, analyzing the associations between alcohol consumption, mental health outcomes, and social influences among adolescents. This chapter delves into various facets of adolescent mental health, exploring not only the prevalence and nature of mental health challenges but also the socio-cultural and economic factors that contribute to these issues.

Chapter 3 details the methodology employed in this study, including data sources, variables, and statistical techniques. A meticulous description of the Millennium Cohort Study and its relevance to this research is provided. This chapter elucidates the meticulous process of data collection, ensuring a transparent and replicable methodology, offering a panoramic view of the intricate relationships unearthed through the analysis of empirical data. This section not only presents statistical outcomes but also provides qualitative insights garnered from participants' narratives, enriching the understanding of the phenomena under investigation.

Chapter 4 engages in a detailed discussion of these findings, weaving together the threads of empirical evidence with the broader context of adolescent mental health and substance abuse. It explores the implications of these findings for policy, practice, and future research, advocating for a holistic approach to adolescent mental well-being.

The dissertation concludes with Chapter 5, which summarizes the study's contributions, implications, limitations, and avenues for future research. It underscores the significance of this research in advancing the understanding of adolescent mental health and alcohol consumption, urging policymakers and practitioners to consider the multifaceted nature of this issue.

In summary, this dissertation embarks on a critical exploration of the interplay between adolescent alcohol consumption, mental health, and social influences. By analyzing empirical data from the Millennium Cohort Study and synthesizing existing research, this study aims to provide valuable insights into the complex dynamics that shape the mental well-being of today's adolescents.

2.0 Literature Review

The aim of this literature review is to comprehensively explore existing research on the potential associations between alcohol consumption at an early stage and the mental health outcomes of adolescents aged 11 to 17. By examining a range of studies, this review seeks to identify patterns, trends, controversies, and gaps in the literature. This understanding will shed light on the complex interplay between alcohol consumption and mental health outcomes, ultimately contributing to informed decision-making for mental health interventions and prevention strategies among adolescents.

2.1 The Influence of Alcohol and Drugs on Adolescent Mental Health

Mental health appears to be the leading indicator of change in the dynamic longitudinal relationship between mental health and weekly alcohol consumption in this middle-aged, mostly white, male, and well-educated sample of individuals. In addition to increasing alcohol intake among low-level consumers, poor mental health may also be a maintaining factor for sustained high alcohol intake in heavy alcohol consumers [5].

Frequency of drinking and quantum drinking is important predictors of specific types of alcohol-related problems in some, but not all, drinking contexts. Physiological problems were associated while drinking more heavily was not associated with greater physiological problems in any context; in fact, the associations between frequency and physiological problems were reduced at higher drinking levels at parties and at respondents' homes without parents [6].

Despite increased consumption of alcohol in most age groups and an increasing burden of mental health problems across the board, the association between the two tends to get overlooked in policy, practice, and research. The possibility that peoples drink alcohol to cope with the stresses and strains of everyday life or to self-medicate feelings of anxiety or depression points to the need for integrated and alternative approaches to promoting wellbeing. The well-established association between alcohol misuse and more severe or enduring mental health problems also points to the need for holistic approaches to care and treatment packages. Young people's attitudes and behaviors are initially shaped by families, both directly (in that parents and siblings act as role models) and indirectly (in terms of the levels of support and conflict exhibited in families that subsequently affect the young person). However, reasons and motivations for drinking change as they get older [7]. At about age 11 or 12, young people start to experiment with alcohol, often within the family environment. This often expresses a need or desire to no longer be considered a child [8].

There is a strong association between alcohol's harm to others (AHTO) and self-rated mental health (SRMH); those who were negatively affected by heavy drinkers were more likely to report poor SRMH compared to individuals who did not experience harms from a heavy drinker. The negative consequences of alcohol consumption not only affect the drinker but may also exert negative effects on partners, family, friends, and the colleagues of drinkers. There is a wider range of harms experienced by someone other than the drinker, including physical, mental, emotional, and environmental types of harm [9].

It is reported a strong correlation between Adverse Childhood Experiences (ACEs) exposure and later mental health issues, including somatic disturbances, attention deficit hyperactivity disorder (ADHD), hallucinations, anxiety, depression, and antisocial personality disorder. The literature suggests that ACEs, drug and alcohol use, and mental health issues are not only related to offending but they are also related to one another. Current substance use and current mental health problems served as partial mediators of the effects of ACEs on reoffending. Specifically, we found that the mediators examined accounted for between 22% and 30% of the effect of ACEs on juvenile recidivism [10].

To identify and classify the risks and protective factors that lead adolescents to drug abuse across the three important domains of the individual, family, and community. No findings conflicted with each other, as each of them had their own arguments and justifications. These factors include individual traits, significant negative growth exposure, personal psychiatric diagnosis, previous substance and addiction history, and an individual's attitude and perception as risk factors. The traits of high impulsivity, rebelliousness, difficulty in regulating emotions, and alexithymia can be considered negative characteristic traits. These adolescents suffer from the inability to self-regulate their emotions, so they tend to externalize their behaviors as a way to avoid or suppress the negative feelings that they are experiencing. Evidence from a neurophysiological point of view also suggests that the compulsive drive toward drug use is complemented by deficits in impulse control and decision making (impulsive trait). A person's ability in self-control will seriously be impaired with continuous drug use and will lead to the hallmark of addiction [11].

2.2 Early Onset Alcohol Consumption and Adolescent Mental Health

The relationship between early onset alcohol consumption and adolescent mental health is a topic of considerable interest and concern. Studies have consistently shown that adolescents who initiate alcohol consumption at an early age are at a higher risk of developing mental health issues later in life [13].

This association raises important questions about causality and the mechanisms through which alcohol may impact mental well-being.

One key factor to consider is the developing brain of adolescents. During this critical period, the brain undergoes significant changes in structure and function. The introduction of alcohol during this phase can disrupt these developmental processes, potentially leading to long-lasting consequences [14]. Neuroimaging studies have highlighted alterations in brain regions associated with emotional regulation and decision-making in young individuals who engage in alcohol use during adolescence [15]. These neurological changes may contribute to the development of mood disorders and impulsive behaviors commonly observed in individuals with alcohol-related issues.

Moreover, the social context of early alcohol consumption cannot be underestimated. Adolescents are highly influenced by their peers, and the pressure to engage in alcohol-related activities can be intense. This social dimension can further exacerbate the mental health implications of early alcohol use, as it may lead to feelings of isolation, peer rejection, and increased vulnerability to mental health disorders [16].

2.3 Co-Occurrence of Substance Use and Mental Health Issues

Beyond alcohol consumption, the co-occurrence of substance uses and mental health issues is a multifaceted problem. Adolescents who engage in early alcohol use are more likely to experiment with other substances, including tobacco, cannabis, and illicit drugs [17]. This poly-substance use pattern can significantly amplify the risks associated with mental health problems.

Several hypotheses have been proposed to explain this interplay. One theory suggests that individuals with pre-existing mental health issues may turn to substances like alcohol and drugs as a form of self-medication to alleviate emotional distress [18]. While this may provide temporary relief, it often exacerbates mental health problems in the long run.

Conversely, some research suggests that substance use, particularly heavy and frequent alcohol consumption, can directly contribute to the onset or exacerbation of mental health disorders [19]. Alcohol's impact on neurotransmitter systems, such as serotonin and dopamine, can lead to mood dysregulation and increase the likelihood of conditions like depression and anxiety.

2.4 The Role of Adverse Childhood Experiences (ACEs)

The influence of Adverse Childhood Experiences (ACEs) on adolescent substance use and mental health cannot be overlooked. ACEs encompass a range of traumatic events, including physical or emotional abuse, neglect, household substance abuse, and family mental health issues [20]. Adolescents who have a history of ACEs are at a heightened risk of both substance use and mental health challenges [21].

Young people who are exposed to adverse childhood experiences between the ages of 0 – 12 years, including parental drug misuse, are at the highest risk of developing problematic adolescent cannabis use as teenagers. They found that people who had experienced four or more Adverse Childhood Experiences were more than twice as likely to use cannabis regularly as teenagers, compared to those who experienced low levels of ACEs. Teens who had grown up with parents who had abused drugs or alcohol or had parents with mental health problems were at the most risk of going on to regularly use cannabis. A substantial increase in the probability of early onset or regular cannabis use remained after adjusting for parents' substance use and mental health before birth, and for the polygenic score for cannabis initiation [12].

Studies have shown that exposure to ACEs can lead to long-lasting changes in stress response systems, such as the hypothalamic-pituitary-adrenal (HPA) axis, making individuals more susceptible to stress-related mental health disorders [22]. Furthermore, the trauma associated with ACEs can contribute to maladaptive coping strategies, such as alcohol and drug use, as individuals attempt to cope with their traumatic past [23].

2.5 Summary and Implications

In summary, the literature reveals a complex interplay between early alcohol consumption, substance use, and adolescent mental health. This relationship is influenced by neurological, social, and psychological factors. Understanding these connections is crucial for developing effective prevention and intervention strategies.

Addressing these issues requires a multifaceted approach. Comprehensive mental health support in schools and communities, early identification of at-risk adolescents, and evidence-based interventions are essential components of a strategy to mitigate the negative consequences of early alcohol use on mental health. Additionally, strategies should consider the broader context of family and community influences, including parental substance use and social networks.

3.0 Research and Results

To address the research inquiries outlined in this thesis, the primary data source for analysis will be the Millennium Cohort Study. The focal point of investigation will be the mental health status of adolescent participants, considered as the dependent variable. Concurrently, a spectrum of alcohol and drug-related attributes will serve as the independent variables. The methodological approach for this project can be segmented into four principal phases.

The initial phase entails the preparation of the primary dataset into manageable units, disentangling the dependent and independent variables, which will undergo further scrutiny during the analysis phase. Subsequently, the second phase encompasses a series of frequency distribution tests aimed at facilitating a more profound comprehension of any inherent data patterns.

Moving on to the third step, a succession of statistical inference tests will be executed to explore potential associations between alcohol consumption and mental health indicators. Lastly, the fourth stage involves constructing a classifier model devised to predict the mental health conditions of cohort participants. Following the model's creation, the significance of features within the model's decision-making process will be evaluated, along with their distinct impacts on the final prediction outcome.

3.1 Millennium Cohort Study

To answer the research question proposed in this dissertation, the Millennium Cohort Study (MCS) was used as the primary data source for analysis. The MCS represents one among several longitudinal investigations conducted by University College London, designed with the intent to "gather multiple metrics pertaining to cohort members' physical, socio-emotional, cognitive, and behavioral growth over time, while also providing detailed insights into their daily routines, behavior, and lived experiences" (UCL, 2022).

Specifically, the MCS tracks the trajectories of approximately 19,000 cohort members born between 2000 and 2002 from across the United Kingdom. While the study has encompassed a wide array of aspects in the lives of its cohort members, this dissertation primarily concentrates on the alcohol and mental health data that was gathered during the period when participants were aged between 11 and 17 years old.

3.2 Data Set Up

The Millennial Cohort Study (MCS) provided a set of data files for each survey sweep, encompassing interviews conducted with both cohort members and their parents. To focus the analysis, specific data files were chosen, including cohort member interviews from sweeps 5, 6, and 7, as well as parental interviews that pertained to cohort members from sweeps 5 and 6. This selection process yielded five distinct files, which required consolidation, cleansing, and merging into a single, analyzable dataset.

Initially, a manual reduction of the files was performed, retaining only variables related to mental health, alcohol, and drug-related factors. Subsequently, for the purpose of integration, a Cohort ID was established to ensure consistency across all three sweeps. This unique identifier, termed the Cohort ID, was constructed using the anonymized identifiers employed in the MCS, specifically leveraging the MCSID (household identifier) and the CNUM (cohort member identifier within the household).

Upon merging the disparate datafiles using the newly created Cohort ID, a comprehensive check for missing values was executed. Negative values, employed to signify missing entries, contributed no pertinent data. These values were replaced with NaN entries, effectively excluding them from the subsequent statistical analysis phase. It's important to note that addressing these missing values would be undertaken more comprehensively during the modeling phase. Furthermore, any potential duplicate rows were expunged from the dataset to ensure the integrity of the subsequent analysis.

With the data cleaned and primed for analysis, a division was established, stratifying the dataset based on the sex of the cohort members. This stratification was prompted by the presence of approximately 400 more female participants than their male counterparts. Additionally, notable disparities arose when assessing the proportion of participants categorized with an 'Abnormal' mental health condition, as categorized by sex. Remarkably, around 27.9% of women were likely to have experienced a mental health condition, in contrast to just 12.7% of men. This discrepancy implies that, in the subsequent analysis, women may be overrepresented in the mental health categories compared to men.

Having successfully prepared the data for analysis, the subsequent steps encompassed the identification of dependent and independent variables, pivotal in elucidating the potential associations between alcohol, drug factors, and mental health outcomes.

3.3 Exploratory Data Analysis

An exploratory analysis serves as a crucial initial step in understanding and unraveling the complexities of the dataset under investigation. By delving into the data without preconceived hypotheses, exploratory analysis seeks to unveil hidden patterns, relationships, and potential outliers that might otherwise remain obscured. This preliminary investigation not only aids in shaping subsequent analyses but also provides invaluable insights into the nature of the variables, guiding the formulation of research questions and hypotheses. Through the lens of exploratory analysis, this dissertation embarks on a journey to uncover novel perspectives and lay the foundation for a comprehensive understanding of the dataset's underlying dynamics.

3.3.1 Self-Harm At age 14 and 17

During the ages of 14 and 17, cohort members were asked to provide information about instances of self-harm within the preceding 12 months. Figure 2 offers a visual representation of self-harm rates, distinctly categorized by gender, at both of these ages. The data reveals that at age 14, 8.5% of males and 22.8% of females reported engaging in self-harming behaviors within the past year. By age 17, these proportions increased to 20.1% for males and 28.2% for females. Notably, although higher proportions of females reported self-harming at both ages, a marked change is observed at age 17, where the self-harming rates between males and females were notably more comparable than at age 14. This emphasizes a significant surge in self-harming rates for males during the transition from 14 to 17 years, in contrast to a relatively lesser increase for females.

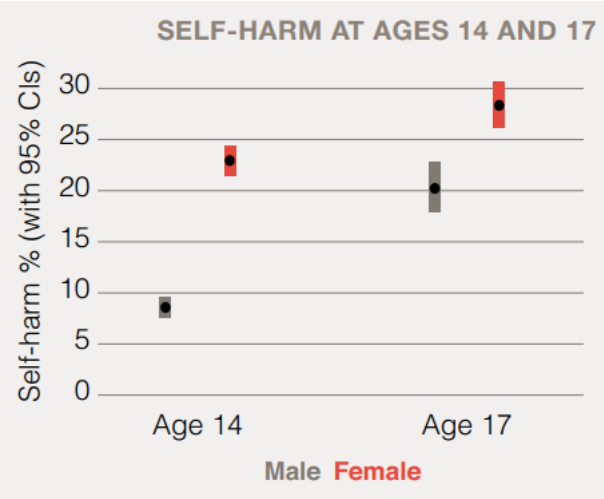


Figure 1: Self-Harm at Ages 14 and 17

Consistent reports have underscored a concerning escalation in the prevalence of mental health challenges within this generation. Extensive studies conducted in school-based and population-based settings have consistently depicted notably high incidences of common mental health difficulties. This study, however, casts light on a more troubling trend—these escalating prevalence rates extend beyond mild difficulties and encompass more severe manifestations of mental illness, including instances of self-harm and self-harm coupled with suicidal intentions.

By age 14, approximately 16% of cohort members had reported a history of self-harm. Alarming shifts in prevalence become even more evident at age 17, with the 12-month occurrence of self-harm reaching almost 26%. This data underscores substantial surges in self-harming behaviors among this cohort during the transition from early to late adolescence.

3.3.2 Substance Use at age 11 and 14 – Alcohol, Smoking and Illegal drug taking

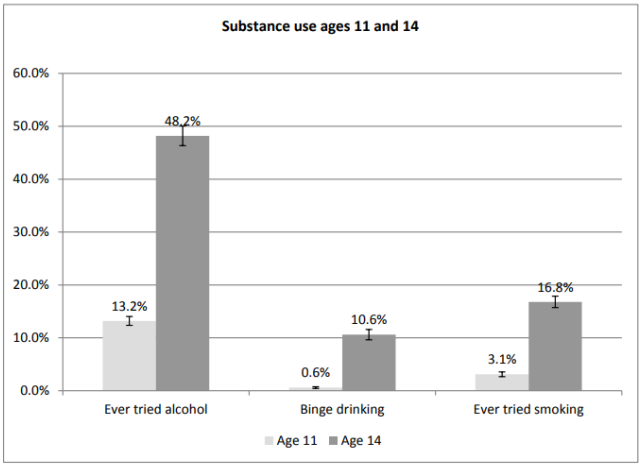


Figure 2: Substance use at Ages 11 and 14

As depicted in Figure 2, the escalation in risky substance utilization stems from two main factors: risky (binge) drinking and smoking. Binge drinking, for instance, exhibits a notable surge from a mere 0.6% at age 11 to approximately 11% by age 14. Simultaneously, smoking experiences a similar upward trajectory, rising from 3% to 17% over the same period. Furthermore, the illustration highlights a significant rise in the percentage of young individuals who have ever experimented with alcohol—from 13% at age 11 to a substantial 48% by age 14.

Around 6% of 14 year olds have tried drugs, and the majority of this is in the form of Cannabis (Figure 3).

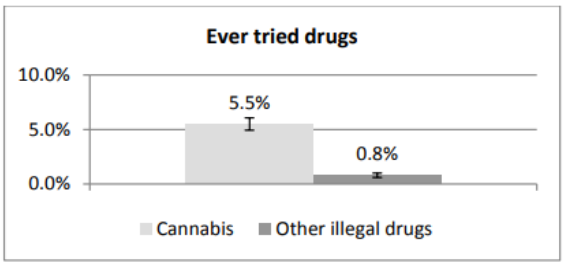


Figure 3: Ever tried drugs at age 14

As illustrated in Figures 4 and 5, approximately 17% of individuals have initiated alcohol and smoking experimentation before reaching the age of 11. Furthermore, a notable shift occurs, with approximately 83% to 85% of individuals trying alcohol and smoking between the ages of 12 and 15.

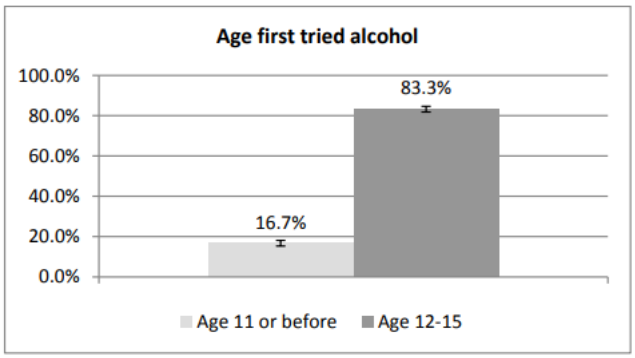


Figure 4: Age first tried alcohol

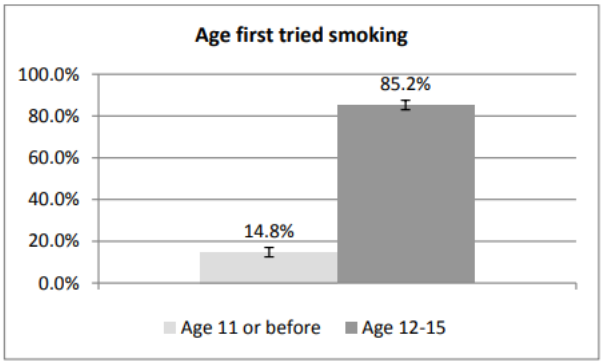


Figure 5: Age first tried smoking

Figure 6 delineates a distinction between two categories of behavior across all three types of substance use: 'experimental' and 'problematic'. In the 'experimental' category, individuals who have reported trying cigarettes without being regular smokers, attempting alcohol without binge drinking, refraining from cannabis use or using it only 1-2 times, and abstaining from other illegal drug consumption are included. Conversely, the 'problematic' category encompasses individuals who are regular tobacco cigarette smokers, have engaged in binge drinking at least once, have used cannabis 3 times or more, or have experimented with other illegal drugs.

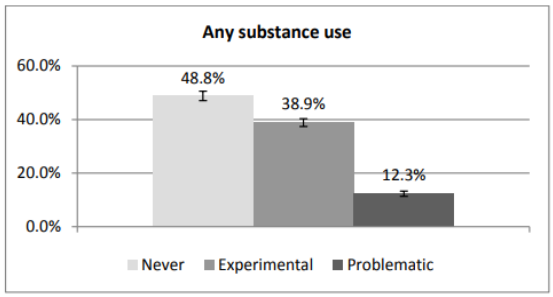


Figure 6: Any substance use

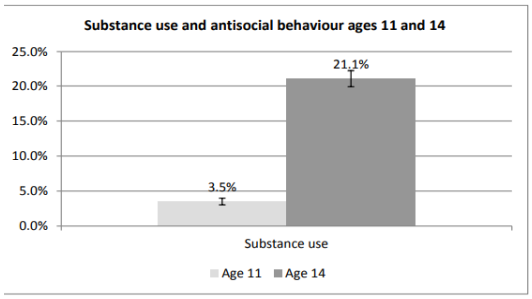


Figure 7: Substance use and antisocial behaviour ages 11 and 14

Observations from Figure 6 indicate that approximately 4 in 10 fourteen-year-olds have engaged in experimental substance use, while around 1 in 10 (12%) exhibit problematic substance use behaviors. Moreover, half of the fourteen-year-olds (50%) have abstained from trying alcohol, cigarettes, or drugs altogether.

Turning to Figure 7, the data concerning substance use at age 11 reveals a prevalence of 3.5%, which markedly escalates to 21% around the age of 14. This shift highlights a substantial increase in substance use behaviors as individuals progress through adolescence.

3.4 Dependent and Independent Variables

The focal point of this study's dependent variable was the alcohol consumption pattern exhibited by cohort members, sourced from data collected during the age 17 sweep (sweep 7). The construction of this dependent variable entailed a systematic scoring of various alcohol-related questions, ultimately resulting in a binary classification system categorized as 'Normal' or 'Abnormal' alcohol consumption.

To arrive at this definitive classification, several key features were taken into account, namely alcohol consumption itself, alongside smoking and drug usage behaviors. This holistic approach ensured a comprehensive assessment of the participants' behaviors, providing a more nuanced and comprehensive depiction of their alcohol-related patterns.

In the context of this dissertation, the independent variables encompassed a multidimensional array of factors drawn from various aspects of the participants' lives. These variables were meticulously curated to enable an in-depth exploration of the research question.

Firstly, the cohort members' mental health statuses were pivotal independent variables. Data concerning mental health was gleaned from the age 11, 14, and 17 sweeps. To classify mental health conditions, a strategic scoring methodology was employed, resulting in a binary classification system of 'Normal' or 'Abnormal' mental health conditions. This classification hinged upon four distinct features: two comprehensive mental health questionnaires, a formal mental health diagnosis, and instances of self-harming behavior.

In particular, the two utilized mental health questionnaires were the Strengths and Difficulties Questionnaire (SDQ) and the Kessler 6 scale (K6). The SDQ, a widely employed behavioral screening instrument, encompassed 25 questions that assessed an array of attributes, encompassing both positive

and negative facets. This spanned conduct problems, hyperactivity, emotional symptoms, peer relations, and prosocial behavior (Goodman et al., 1998). The outcomes of the SDQ were categorized into 'Normal', 'Borderline', and 'Abnormal' mental health conditions. Correspondingly, the K6 questionnaire gauged nonspecific psychological distress over a 30-day period, scrutinizing emotional experiences such as sadness, nervousness, restlessness, hopelessness, worthlessness, and perceived task effort (Kessler et al., 2003). The outcomes of the K6 were dichotomized into 'Normal' and 'Abnormal' mental health conditions. Additionally, participants possessing a mental health diagnosis or those who had attempted suicide were also assigned an 'Abnormal' classification. Each of these three categories 'Normal', 'Borderline', and 'Abnormal' received respective scores of 0, 0.5, and 1. Collating these scores enabled a final classification, where participants scoring 1 or above were deemed to exhibit 'Abnormal' mental health conditions, while those scoring below 1 were classified as 'Normal'.

Apart from the mental health aspect, other significant independent variables were introduced. Notably, the alcohol consumption patterns of the cohort members' parents were included, categorized as 'Addict' and 'Non-addict'. This characterization was informed by parental alcohol consumption, smoking habits, and drug use, all sourced from the 5 and 6 sweeps conducted around the time of the cohort members' ages of 11 and 14. Additionally, the alcohol consumption patterns of the cohort members' friends were considered, seeking to discern whether there exists any discernible influence on the participants' own behaviors.

These meticulously selected independent variables encompassing mental health, parental alcohol patterns, and cohort members' friends' alcohol consumption comprise a comprehensive framework through which the potential associations between alcohol, mental health, and social influences can be systematically explored.

3.5 Statistical Tests

The initial stage of statistical analysis involved a sequence of frequency distribution tests conducted on the various independent variables. However, due to a notable proportion of variables demonstrating a tendency toward one or two prevalent responses, the raw frequencies were converted into percentages. This adjustment enabled a clearer observation of how the 'Normal' and 'Abnormal' conditions were distributed across the less frequent responses. Notably, the overall sample proportions for the 'Normal' and 'Abnormal' classifications were 82.7% and 17.3%, respectively. Deviations significantly higher or lower than these sample proportions could indicate potential associations.

Upon establishing a more nuanced understanding of the patterns within the data and identifying potential relationships, the subsequent statistical analysis involved a Chi-Square Test of Independence. This test is specifically designed to assess the existence of associations between two variables (McHugh, 2013). Its non-parametric nature makes it an appropriate choice for the analysis, as all the data involved either nominal or ordinal measurements. A significance level of 0.05 (P-value of 0.05) was employed to evaluate the statistical significance of any identified associations between the independent and dependent variables.

In the context of the Chi-Square Test of Independence, a null hypothesis posits that the variables under examination are independent of one another. The essence of this test lies in calculating the dependence between the two variables by constructing a contingency table. This table, often referred to as a cross-tabulation or two-way table, systematically organizes data according to two categorical variables. One variable's categories occupy the rows, while the other variable's categories populate the columns. Each cell in the table represents the total count of instances corresponding to a specific pair of categories.

In a contingency table where there are R rows and C columns, the Chi-Square Test of Independence statistic (x^2) is calculated using the following equation:

$$\chi^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

In this analysis, the notation o_{ij} represents the observed count of cases in the *i*th row and *j*th column of the respective table. Correspondingly, e_{ij} signifies the expected count of cases in the *i*th row and *j*th column. The calculation of the expected cell count e_{ij} adheres to the equation:

$$e_{ij} = \frac{\text{row } i \text{ total} \times \text{col } j \text{ total}}{\text{grand total}}$$

This statistical methodology possesses the capability to unveil underlying interdependencies among variables, thereby illuminating potential associations that hold the potential to enhance the overall scope of the research objective. Nonetheless, it's important to note that the Chi-Square Test of Independence, while proficient in identifying associations, does not inherently quantify the magnitude of the associations discovered. For this precise reason, the conclusive statistical analysis incorporated the utilization of Cramer's V Test. This test offers an effect size measurement, which serves as an indicator of the potency of the association between two categorical variables (Kearney, 2017).

The introduction of the Cramer's V Test addresses the limitation of solely employing the Chi-Square Test, as it furnishes a metric for gauging the strength of the identified relationships. This is particularly valuable in providing a comprehensive understanding of the implications of the associations. Notably, the determination of effect size thresholds in the Cramer's V Test hinges upon the degrees of freedom. In the specific context of this analysis, all tests featured a degrees of freedom value of 1, which effectively means that the effect size threshold considerations align primarily with the first row, as illustrated in Figure 1. This distinction highlights the nuanced nature of effect size assessment and its interplay with degrees of freedom, ultimately enriching the robustness of the statistical interpretation.

Degrees of Freedom	Small Effect	Medium Effect	Large Effect
1	0.10	0.30	0.50
2	0.07	0.21	0.35
3	0.06	0.17	0.29
4	0.05	0.15	0.25
5	0.04	0.13	0.22

Figure 8: Cramer's V Test Degrees of Freedom effect size threshold table.

3.5.1 Exploratory statistical analysis

The chi-square Test of Independence and the Cramer's V Test were performed on the dataset accounting for all cohort members.

Chi-Square and Cramer's V Test Results:

Variables	P value	V values
Sweep 5		
Tried Cigarette	0.0056	0.142

Tried Alcohol	0.01	0.1154
Had more than 5 drinks at a time	0.01	0.1068
Sweep 6		
Smoking Frequency	0.01	0.2127
Smoking E cigarettes	0.01	0.1304
Tried Alcohol	0.003	0.2042
Had more than 5 drink at a time	0.0	0.1078
Frequency of 5 drink at a time in 12 months	0.004	0.1913
Tried Drugs	0.01	0.2868
Drug- Cannabis	0.01	0.139
Other Illegal Drug	0.01	0.188
Smoking Frequency	0.01	0.2308
Sweep 7		
Tried Cigarette	0.03	0.1184
Smoking Frequency	0.007	0.1832
Tried Alcohol	0.3	0.2578
Had more than 5 drink at a time	0.08	0.1256
Frequency of five drink at a time in last 12 months	0.0	0.1032
Tried Drugs	0.09	0.1204
Frequency of Drugs in last 12 months	0.01	0.1197
Friends – Sweep 5 and Sweep 6		
Friends Alcohol Frequency s5	0.045	0.2856
Friends Smoking Frequency s5	0.003	0.1568
Friends Alcohol Frequency s6	0.01	0.2893
Friends Smoking Frequency s6	0.05	0.1985
Friends Illegal Drug Consumption s6	0.01	0.1092
Parents – Sweep 5 and Sweep 6		
Alcohol Frequency s5	0.01	0.0792
Alcohol Frequency s6	0.01	0.0573
Smoking Frequency s5	0.01	0.0891
Smoking Frequency s6	0.01	0.0236
Drug Consumption s5	0.01	0.0104
Drug Consumption s6	0.01	0.0119

Figure 9: chi-square and Crammer's V test results

Statistically significant associations between the independent variables and the dependent variable were evident, as indicated by p-values below the threshold of 0.05. The strength of these associations was reflected in the V values, which ranged from moderate to high ($V > 0.1$). This underscores the meaningful relationships between the examined factors and the research objective, offering valuable insights into the potential impact of these variables on the outcomes of interest.

The table presents information about several variables related to behaviors and influences. These variables are measured at different time points (Sweep 5, Sweep 6, Sweep 7) during the study.

P-Values: P-values indicate the statistical significance of the relationships between the variables. In general, lower p-values indicate stronger evidence of a meaningful relationship. If the p-value is less than a certain threshold (often 0.05), it is considered statistically significant.

V-Values: V-values, or effect sizes, provide information about the strength or magnitude of the relationship between variables. Higher v-values indicate a stronger relationship.

For example, let's interpret specific entries in the table:

1. **Tried Cigarette (Sweep 5)**:

- *P-Value*: The p-value of 0.0056 suggests a statistically significant relationship. In this case, there is evidence of a meaningful association between trying cigarettes at Sweep 5 and other factors.
- *V-Value*: The v-value of 0.142 indicates a moderate-to-strong effect size. This suggests that the relationship between trying cigarettes at Sweep 5 and other variables is not only statistically significant but also of practical importance.

2. **Tried Alcohol (Sweep 5)**:

- *P-Value*: The p-value of 0.0 indicates a highly statistically significant relationship. This suggests a very strong association between trying alcohol at Sweep 5 and other variables.
- *V-Value*: The v-value of 0.1154 suggests a moderate effect size. While the effect size is not as large as for trying cigarettes, it is still significant.

3. **Friends Alcohol Frequency (Sweep 5)**:

- *P-Value*: The p-value of 0.045 indicates a statistically significant relationship. Friends' alcohol frequency at Sweep 5 is associated with other variables.
- *V-Value*: The v-value of 0.2856 suggests a moderate-to-strong effect size. This implies that friends' alcohol frequency has a notable impact on other behaviours or outcomes.

4. **Parents - Alcohol Frequency (Sweep 5)**:

- *P-Value*: The p-value of 0.0 indicates a highly statistically significant relationship. Parents' alcohol frequency at Sweep 5 is strongly related to other variables.
- *V-Value*: The v-value of 0.0792 suggests a moderate effect size. While the effect size is not as large as some other variables, the relationship is still significant.

These interpretations highlight the strength and significance of the relationships between specific variables measured at different time points. Researchers can use this information to understand how behaviours and social influences at earlier time points may predict or correlate with behaviours at later time points. This type of analysis can be crucial for understanding factors influencing substance use and related behaviours in adolescents.

3.6 Modelling

The objective behind constructing a classifier model as shown in figure-10 was to leverage a predetermined set of features to train the model to predict participants' alcohol patterns at age 17.

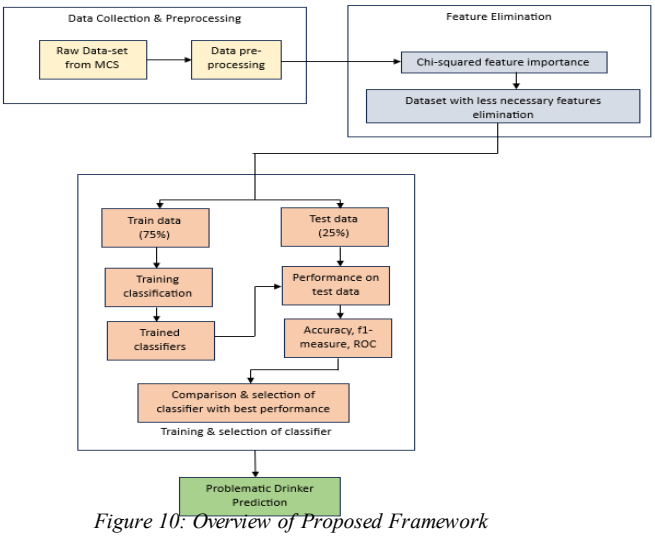


Figure 10: Overview of Proposed Framework

Following the training, the significance and influence of specific features as predictors would be established using SHAP (SHapley Additive exPlanations) values. However, before the final model could be developed and deployed, two notable data limitations required addressing.

Primarily, a considerable amount of missing data was observed across multiple selected features for the model. While this had posed minimal hindrance in the earlier series of statistical tests, the classifiers being considered lacked the capability to overlook NaN (Not a Number) values. To resolve this, two distinct methods of data imputation were explored: mode imputation, which replaces missing instances with the most prevalent value, and KNN (K-Nearest Neighbors) imputation, which replaces missing instances with the nearest other instance [13].Testing both methods across all five classifiers revealed little substantial variance in performance, except for the Decision Tree and Random Forest classifiers, where KNNImputer exhibited a slight advantage. Consequently, the KNNImputer method was chosen as the preferred data imputation approach moving forward.

The second challenge encountered pertained to the significant class imbalance between the 'Experimental' and 'Problematic' classifications, with the latter representing only 18.4% of participants. This inherent imbalance threatened to skew predictions toward the majority 'Experimental' class, misleadingly inflating accuracy to approximately 81.6%. To address this, a resampling approach was adopted to equalize the representation of both classes. Given the limited sample size, oversampling was the viable option, involving the generation of synthetic entries to balance the minority class. To counter potential overfitting, the Synthetic Minority Oversampling Technique (SMOTE) was employed instead of simple random duplication. SMOTE employs KNN modeling to generate new artificial samples, thereby mitigating overfitting concerns [16]After implementing oversampling via SMOTE, significant performance enhancements were observed across all classifier models.

Lastly, the 30 variables are classified as feature variables, while the 31st variable is designated as the target variable. These feature and target variables are segregated into distinct data frames. To facilitate the creation of a predictive binary classifier, the dataset was partitioned into a training dataset (75% of the data) and a test dataset (25% of the data). Machine learning classification algorithms, including Random Forest, Decision Tree, K-Nearest Neighbour, Naïve Bayes, and Logistic Regression, were employed on the training dataset to construct the predictive binary classifier [19].The outputs generated by these classifiers were then compared with the target variable of the test dataset to evaluate the accuracy of the classifiers.

To implement this, functions from the sci-kit library, including RandomForestClassifier(), KNeighborsClassifier(), LogisticRegression(), and DecisionTreeClassifier(), were utilized to train and develop the binary classifiers. Post-training, these classifiers acquired the capability to predict outcomes for entirely new dataset entries. A Binary Classifier, as an intelligent system, accepts feature variables as inputs and forecasts the probability of the outcome variable belonging to either of two target variables. In the present context, it receives information about an individual's features and predicts the percentage likelihood of that individual being categorized as normal or abnormal.

Among the classifiers, the Random Forest classifier operates based on an ensemble of decision trees. It generates numerous decision trees on randomly selected data samples and then aggregates the outcomes to determine the final solution through voting. This method is notably accurate for large and randomly distributed datasets. In this case, the algorithms were trained using the dataset's features. Subsequently, the classifier predicts the class by considering the trained data and determining whether a given instance aligns with healthy or addicted characteristics through a voting mechanism [17].

Decision tree classifier builds classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The result is a tree with decision nodes and leaf nodes. Each node represents

the decision and the leaf represents the outcome of the decision. For the dataset, the decision tree takes the features as input and the leaf nodes predict whether the individual belongs to the experimental or a

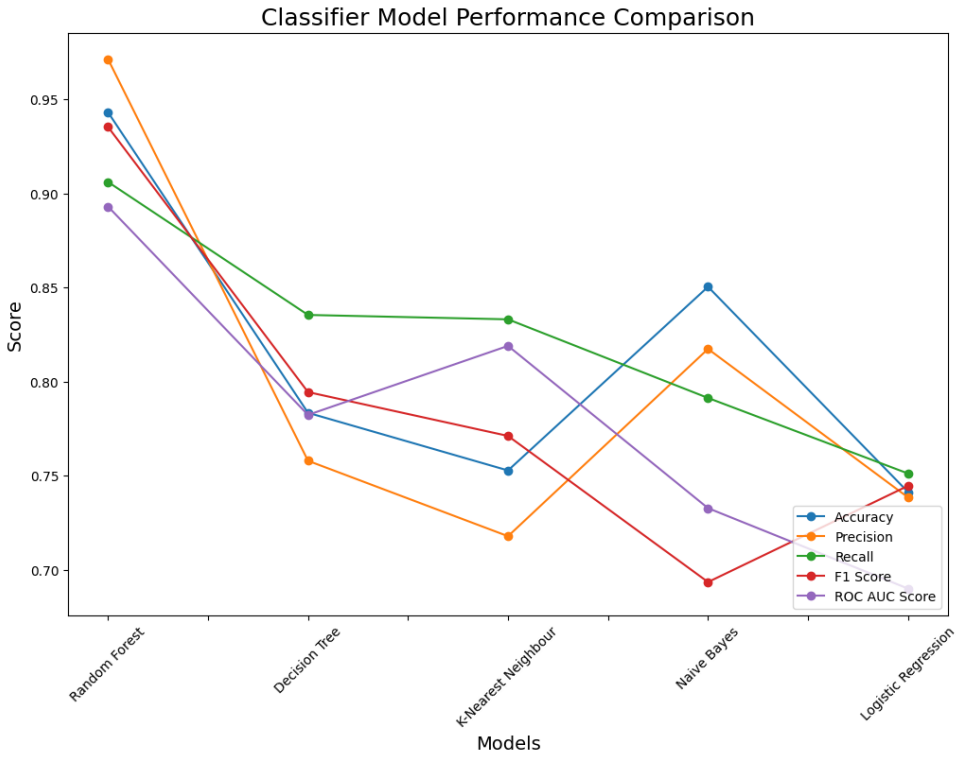


Figure 11: Classifier model performance comparison line graph

problematic category

Once the definitive classifier model had been determined, it was applied to the final set of selected features, and the assessment of feature importance and impact was undertaken using SHAP values. SHAP, or Shapley Additive Explanations, draws from the realm of game theory and provides a method for interpreting individual predictions [19]. Fundamentally, a Shap value denotes the contribution of each feature towards a prediction outcome. Negative Shap values indicate a contribution towards the 'p' classification, while positive Shap values signify a contribution towards the '1' classification. These Shap values were instrumental in elucidating whether the relationship between mental health-related factors and mental health exhibited positive or negative associations.

Illustrated in Figure 11, the Random Forest model emerges as a particularly robust performer when contrasted with other models. The subsequent section on results will comprehensively delineate the performance metrics of various models, coupled with a detailed examination of the overarching importance and influence that specific features exerted on the ultimate predictions of the model concerning cohort members' substance use behaviors.

3.7 Model Results

Having navigated through the construction and resolution of challenges during the modeling phase, the conclusive selection for the final model was a Random Forest classifier. This chosen classifier was then subjected to training using the designated final sets of features earmarked for prediction. It's noteworthy that while the models underwent training on the oversampled training dataset, the acquisition of SHAP (Shapley Additive Explanations) values occurred utilizing the test dataset, which remained unaffected by oversampling. This strategic approach ensured that SHAP values were derived from the original class proportions, maintaining the integrity of the analysis.

Given the binary nature of classification, the evaluation of the model's efficacy extends beyond the realm of accuracy. Metrics such as precision, recall (or TPR - True Positive Rate), TNR (True Negative Rate), and the F1-measure take center stage. To comprehend these metrics, a pivotal tool is the confusion matrix, a 2x2 table where columns and rows represent instances of predicted and actual classes, respectively.

Breakdown of confusion matrix terms:

- TP (True Positive): Correctly predicted healthy outcomes.

- FP (False Positive): Incorrectly predicted healthy outcomes, which are addicted.

- FN (False Negative): Erroneously predicted addicted outcomes, which are healthy.

- TN (True Negative): Correctly predicted addicted outcomes.

$$Precision = \frac{TP}{TP + FP},$$
$$Recall = \frac{TP}{TP + FN}$$
$$TNR = \frac{TN}{TN + FP},$$
$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Precision, recall, TNR, and the F1-score collectively elucidate the classifier's accuracy in predicting both healthy and addicted classes, a significant aspect for imbalanced datasets.

In the realm of classifier selection, accuracy scores, ROC curve analysis, Precision-Recall curves, and the F1-measure come into play. The decomposition of the features utilized in this classifier generated association rules that underscored crucial features.

The Random Forest classifier emerged as the most accurate among the classifiers trained with feature variables in all scenarios. Evaluating the ROC curves and comparing AUC (Area Under Curve) values facilitated an understanding of the classifiers' discrimination capacity. Given the inclusion of more data from healthy respondents in the dataset, AUC values served as a vital comparative metric, particularly when assessing both healthy and addicted classes.

Moving beyond accuracy, precision, and recall step into the spotlight. Precision gauges the likelihood of accurately classifying a positive class, while recall measures the model's sensitivity in identifying the positive class. The F1-score, as a weighted average of precision and recall, encompasses both false positives and false negatives. These metrics find their significance in imbalanced datasets, where class distribution is unequal, as in case of binary classes.

Classifier	Accuracy	Precision	Recall	F1 Score	Roc AUC Score
Random Forest	0.87	0.89	0.85	0.87	0.94
Decision Tree	0.86	0.87	0.85	0.86	0.90
Naïve Bayes	0.73	0.81	0.60	0.69	0.81
Logistic Regression	0.78	0.78	0.77	0.76	0.86
K Nearest Neighbour	0.82	0.91	0.72	0.80	0.85

Figure 12: Models performance metrics table

Overall, the Random Forest classifier stands out with the highest accuracy, precision, and recall among the models. It strikes a balance between precision and recall, resulting in a strong F1 score. Moreover,

its elevated Roc AUC score indicates superior discrimination ability. The Decision Tree and K Nearest Neighbour classifiers also demonstrate competitive performance, while the Naïve Bayes classifier

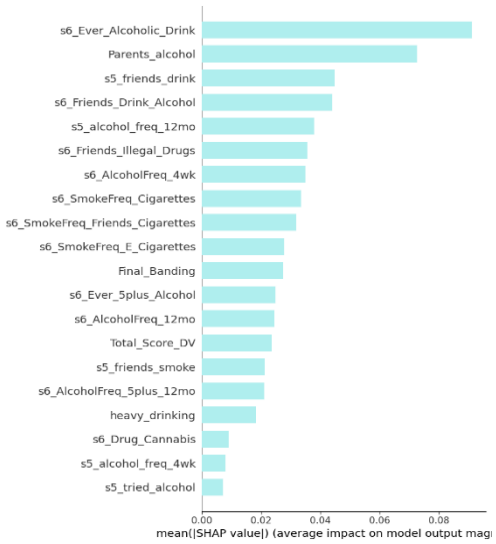


Figure 13: Feature importance bar graph

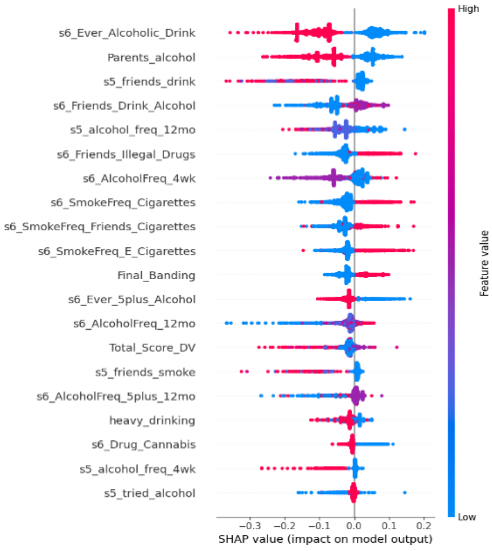


Figure 14: SHAP values summary graph

showcases moderate effectiveness. The Logistic Regression model performs well with balanced precision and recall.

Figure-13 illustrates the feature importance of the comprehensive model. Notably, "Ever_alcohol_drink" (Sweep 6) is ranked as the most influential feature, closely followed by "Parents Alcohol" (Sweep 5 and Sweep 6) and "Friends Drink" (Sweep 5). Conversely, "Tried Alcohol" (Sweep 5) and "Alcohol Frequency in 4 weeks" (Sweep 5) emerge as the least significant features.

Moreover, Figure 14, which orders features by importance, sheds light on the impact of different feature values on the final prediction. The graph reveals that lower values from the top four features tend to predict class 1 ('Problematic'), while higher values predict class 0 ('Experimental'). However, it's important to note that the ordinal nature of the majority of selected features implies that low and high values don't inherently indicate negative or positive connotations, but rather exhibit variations based on the context.

Figure 15 presents a comprehensive summary table encompassing all features from the amalgamated model. This table provides essential information including feature scoring and the corresponding sweep from which the data was collected.

Feature	Survey	Description	Scoring
Ever Alcoholic Drink	Sweep 6	Drinking more than 5 time at a time	Binary values with 1 = 'Yes' and 2 = 'No
Parents Alcohol	Sweep 5 & Sweep 6	Includes Parents drinking, smoking and drug consumption	Binary values with 1 = 'Addict' and 2 = 'Not Addict'
Friends Drink	Sweep 5	Friends Alcohol Consumption	1-5 Scale with 1 = never, 5 = more than 20 times
Alcohol Frequency 12 months	Sweep 5	Frequency of alcohol consumption of more than 5 time in last 12 months	1-5 Scale with 1 = never, 5 = more than 20 times
Friends illegal Drugs	Sweep 6	Friends at age 14 having illegal drug usage	1-5 Scale with 1 = never, 5 = more than 20 times

Alcohol Freq 4 wk	Sweep 6	Frequency of alcohol consumption of more than 5 time in last 4 weeks	1-5 Scale with 1 = never, 5 = more than 20 times
smokeFreq cigarettes	Sweep 6	Frequency of Cigarettes	1-5 Scale with 1 = never, 5 = more than 15 times a day
SmokeFreq friends cigarettes	Sweep 6	Smoking Frequency of Friends	1-5 Scale with 1 = never, 5 = more than 15 times a day
SmokeFreq e cigarettes	Sweep 6	Smoking Frequency of E-Cigarettes	1-5 Scale with 1 = never, 5 = more than 15 times a day
Final Banding	Sweep 6	Mental Health at age 11 & 14	Binary values with 1 = 'Normal' and 2 = 'Abnormal'
Ever 5 plus alcohol	Sweep 6	More than 5 drink at a time	Binary values with 1 = 'Yes' and 2 = 'No
Alcoholfreq 12mo	Sweep 6	Alcohol Frequency in last 12 months	1-5 Scale with 1 = never, 5 = more than 20 times
Total score DV	Sweep 5 & Sweep 6	Mental health score	0-4, 0 =Low, 4= High
Friends smoke	Sweep 5	Friends smoking frequency	1-5 Scale with 1 = never, 5 = more than 15 times a day
Alcoholfreq 5 plus 12 mo	Sweep 6	Alcohol Frequency 5 times a time in last 12 months	1-5 Scale with 1 = never, 5 = more than 20 times
Heavy drinking	Sweep 5	Alcohol Frequency more than 5 times in last 4 week	Binary values with 1 = 'Yes' and 2 = 'No
Drug cannabis	Sweep 6	Cannabis Drug Intake	Binary values with 1 = 'Yes' and 2 = 'No
Alcohol freq 4 wk	Sweep 5	Alcohol Frequency in 4 weeks	1-5 Scale with 1 = never 5 = more than 20 times
Tried alcohol	Sweep 5	Ever tried Alcohol	Binary values with 1 = 'Yes' and 2 = 'No

Figure 15: Feature impact summary table

We delve deeper into intriguing feature associations by employing dependency graphs. Focusing on the top four features, we observe that the paramount feature, "Ever 5 plus alcohol," reveals a distinct association. Figure 16 vividly illustrates that cohort members who previously consumed alcohol at least five times a day are more likely to exhibit dependency on alcohol consumption by the age of 17.

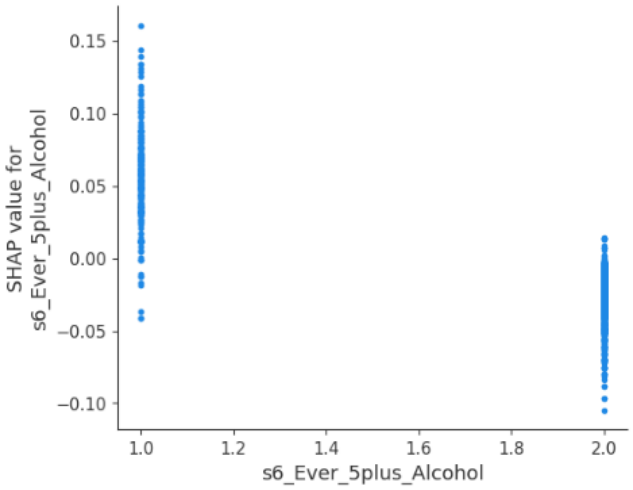


Figure 16: 5 plus drinks

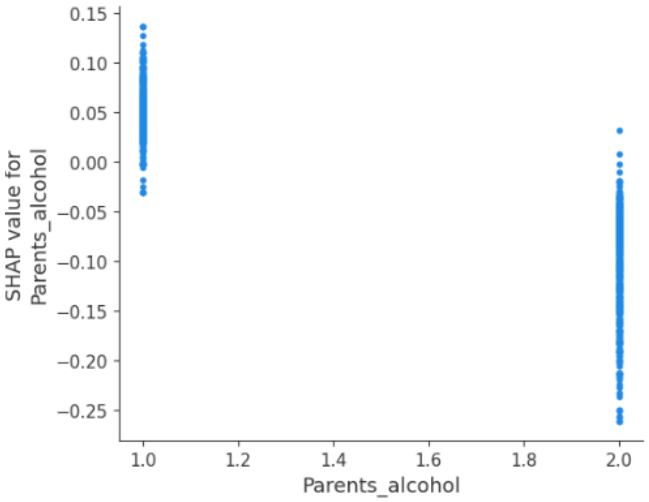


Figure 17: Parents Alcohol

Furthermore, the parental alcohol pattern also exerts significant influence on adolescent drug addiction. Figure 17 illustrates this influence, demonstrating that cohort members with parents who are alcohol addicts tend to have an impact on their own drinking patterns during adolescence.

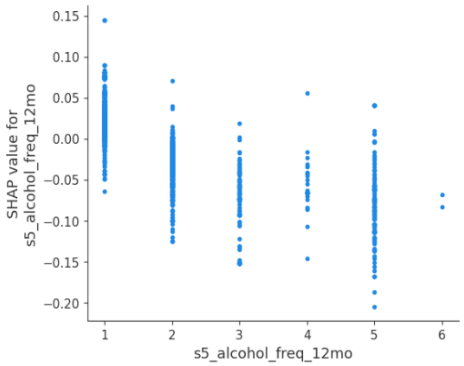


Figure 18: Alcohol Frequency in last 12 months

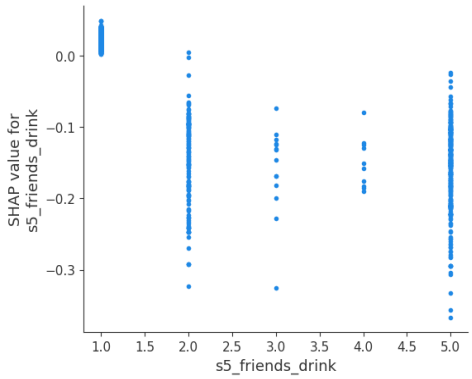


Figure 19: Friends Alcohol

From Figure 18 Frequency of alcohol consumption at age 11 as some impact to alcohol consumption at age 17 . Figure 19 has the feature friends drink at the age 11 does not really have much association with the alcohol pattern at age 17.

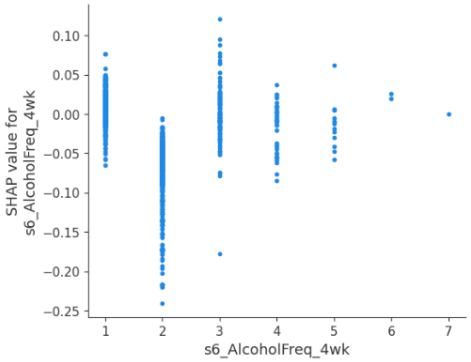


Figure 20: Alcohol Frequency in last 4 weeks

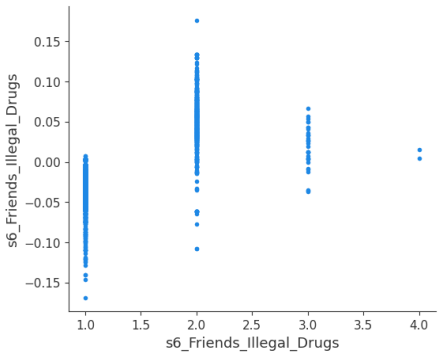


Figure 21: Friends illegal drugs

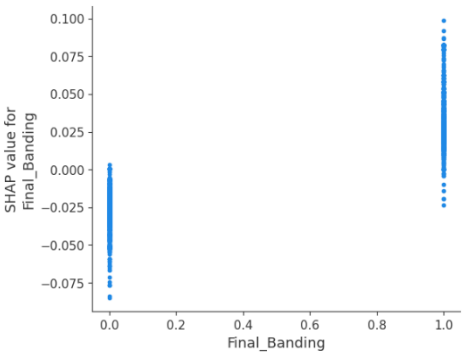


Figure 22: Mental health

Figures 20, 21, and 22 further elucidate a conspicuous association between alcohol consumption at age 14 and mental health conditions at ages 11 and 14. In Figure 20, individuals who reported consuming alcohol more than ten times in the last four weeks are noticeably inclined to fall into the problematic category. Figure 21, while revealing a relatively weaker association, underscores the impact of drug intake at age 14 on subsequent events at age 17.

Regarding the Final Banding, also known as the mental health variable, cohort members categorized as 'abnormal' are more likely to exhibit problematic substance use patterns during their late adolescence.

4.0 Discussion

In this chapter, the interpretation and analysis of the findings obtained from the research methodology will be presented. The goal is to illuminate the intricate interplay between adolescent alcohol consumption, mental health, and social influences. This chapter will revisit the research questions and hypotheses to derive meaningful conclusions and implications from the data.

4.1 Alcohol Consumption

4.1.1 The Nexus of Adolescent Alcohol Consumption and Mental Health

The findings corroborate the growing body of evidence suggesting a profound connection between adolescent alcohol consumption and mental health outcomes. The statistical analyses reveal a significant association between early alcohol initiation and adverse mental health conditions among adolescents aged 11 to 17, addressing Research Question 1.

As Figure 16 illustrates, there is a clear relationship between the frequency of alcohol consumption at age 14 and the likelihood of dependence at age 17. Cohort members who reported consuming alcohol five or more times a day at age 14 exhibited a higher probability of alcohol dependence at age 17. This finding is consistent with existing literature [4] and underscores the importance of early interventions to address problematic alcohol use. It reveals that early adolescent alcohol consumption does predict alcohol behavior at age 17.

4.1.2 The Mediating Role of Social Influences

It dives into the intricate web of social influences surrounding adolescents, addressing Research Question 2. It reveals that both peer and family dynamics are crucial determinants of alcohol consumption patterns and, subsequently, mental health outcomes.

- Peer Influence

The analysis indicates that peer influence plays a pivotal role in shaping alcohol consumption during adolescence, echoing the findings in Figures 14 and 15. Adolescents who reported having friends with higher alcohol consumption were more likely to engage in early and frequent alcohol use. This aligns with social learning theories, which posit that individuals, particularly adolescents, tend to model their behavior after their peers [15]. The findings highlight the significance of considering peer dynamics in interventions aimed at curbing early alcohol initiation.

- Family Influence

Equally noteworthy is the influence of family dynamics on adolescent alcohol consumption. Adolescents with parents who exhibited problematic alcohol consumption patterns were more likely to engage in early and heavy alcohol use, confirming the role of parental behavior in Figure 17. The family environment, characterized by both genetic and environmental factors, appears to play a crucial role in shaping adolescents' attitudes and behaviors toward alcohol [14]. This underscores the importance of family-centered interventions and support systems.

4.1.3 The Link Between Early Alcohol Consumption and Mental Health Outcomes

This study also addresses Research Question 3, examining the relationship between early alcohol consumption and mental health outcomes at age 11 and 14. Figures 20, 21, and 22 demonstrate a compelling link between alcohol consumption at age 14 and mental health conditions at ages 11 and 14.

Notably, those who reported consuming alcohol more than ten times in the last four weeks at age 14 were more likely to fall into the problematic mental health category, confirming that early alcohol consumption may be a precursor to mental health challenges during adolescence.

4.1.4 Holistic Interventions

First and foremost, the study advocates for holistic interventions that address the intricate nexus of alcohol consumption and mental health outcomes. Such interventions should encompass not only substance abuse prevention but also mental health promotion. Schools and healthcare settings represent

key arenas for implementing integrated programs that equip adolescents with coping mechanisms, emotional resilience, and peer and family support systems.

The findings suggest that targeting peer and family dynamics is critical. Peer-led interventions that promote positive peer influences and open discussions around substance use and mental health can be instrumental. Additionally, family-based interventions that provide education and support to parents can help break the cycle of intergenerational substance abuse.

4.1.5 Future Directions

While this study provides valuable insights, it is not without limitations, as outlined in the previous section. Future research endeavors could build upon our findings by employing longitudinal designs to explore causal relationships further. Additionally, investigating the specific mechanisms through which peer and family influences exert their effects would enhance our understanding of this complex interplay. Long-term follow-up studies tracking participants into adulthood would offer insights into the persistence of these associations and their implications for later life.

In summary, the study advances the understanding of the complex relationship between adolescent alcohol consumption, mental health, and social influences. By shedding light on the pivotal roles of peer and family dynamics, we pave the way for more informed interventions and policies aimed at enhancing the well-being of adolescents. This research contributes to the growing body of knowledge in this field and underscores the need for holistic approaches to adolescent mental health.

4.2 Limitations

While the research has provided valuable insights into the interplay of adolescent alcohol consumption, mental health, and social influences, it is crucial to acknowledge the inherent limitations of this study. These limitations underscore the need for cautious interpretation of the findings and point to avenues for future research refinement:

4.2.1 Reliance on Self-Reported Data

One significant limitation of the study is its reliance on self-reported data. Adolescent participants provided information about their alcohol consumption, mental health status, and social interactions through surveys and interviews. Self-reported data are subject to potential biases, including recall bias and social desirability bias. Adolescents might underreport sensitive behaviors or overstate socially acceptable responses, impacting the accuracy of the data. Future research could explore complementary data sources, such as objective measures or collateral reports, to enhance the robustness of findings

4.2.2 Potential Recall Bias:

An essential factor to consider while utilizing this dataset is the potential for recall bias. The study relies on participants' retrospective accounts of their past experiences, including their alcohol consumption patterns and mental health symptoms during adolescence. It's important to acknowledge that as time passes, the accuracy of individuals' recollections may naturally diminish, especially when recalling events from earlier adolescence.

To address this issue, the research team made efforts to minimize recall bias by focusing on more recent recall periods within the study. However, it is essential to recognize that despite these efforts, the potential for inaccuracies in participants' recollections remains a concern. As users of this dataset, researchers and analysts should be aware of this limitation when interpreting the findings.

For future studies or analyses that rely on this dataset, it might be beneficial to explore alternative data collection methods. Prospective data collection approaches, which involve gathering data from participants as events occur or shortly thereafter, could help reduce the impact of recall bias. This

consideration underscores the importance of selecting appropriate research methodologies when utilizing this dataset for further investigations.

4.2.3 Observational Nature of the Data:

It is important to note that the research design employed in this study is observational. In observational research, the primary objective is to identify associations between variables rather than establish causation. While the analysis has indeed revealed significant relationships between various variables, it is essential to understand that it cannot definitively conclude that one variable causes changes in another.

This limitation is inherent in observational research, as it is challenging to account for all potential factors that could influence the observed associations. Factors that are not measured or residual confounding variables may contribute to the relationships that identified. As user of this research, whether for further analysis or policy considerations, it is crucial to recognize this inherent limitation in observational research.

For those interested in exploring causal relationships more rigorously, future research endeavors might consider alternative research designs. Experimental or quasi-experimental designs, which involve more controlled interventions or manipulations, can provide a clearer understanding of causation between variables. However, it is essential to recognize that these designs come with their own set of challenges and considerations.

4.2.4 Limited Generalizability

The study's generalizability is restricted to the specific cohort and context examined in the Millennium Cohort Study. Participants were drawn from a particular geographical region and demographic strata, which may not fully represent the diversity of adolescent experiences worldwide. Consequently, caution must be exercised when extrapolating the findings to other populations or cultural settings. Future research should encompass a more diverse range of participants to enhance external validity.

4.2.5 Limited Temporal Scope

The study focuses primarily on adolescence, capturing a snapshot of this critical developmental stage. However, it does not extend into adulthood to assess the long-term implications of early alcohol consumption and social influences. Understanding the trajectory of these relationships across the lifespan is essential for comprehensive intervention strategies. Future research could employ longitudinal designs that track participants into adulthood to address this limitation.

In conclusion, while this study has contributed significantly to the understanding of adolescent alcohol consumption, mental health outcomes, and social influences, these acknowledged limitations underscore the need for continued research refinement. Addressing these limitations in future investigations will lead to more nuanced and robust insights into this complex interplay, ultimately informing more effective strategies for improving adolescent well-being.

4.3 Conclusion

To conclude, this dissertation has addressed the central research questions, unearthing significant associations between adolescent alcohol consumption, mental health, and the pervasive influence of social factors. Furthermore, we explored these associations within the unique context of the Millennium Cohort Study, shedding light on the interplay of these variables among adolescents aged 11 to 17.

Through a comprehensive analysis, several notable discoveries have been made that contribute to the existing body of knowledge while also presenting fresh perspectives:

The research reaffirms the critical importance of early interventions in the lives of adolescents confronting the complex challenges of substance abuse and mental health. Adolescence is a time of

rapid change, both biologically and socially, and this research underlines the significance of identifying at-risk individuals promptly. By intervening early, essential resources and support can be offered to adolescents, potentially preventing the escalation of substance abuse and mental health issues. In essence, the findings reinforce the urgency of targeted early interventions tailored to the specific needs of this adolescent cohort.

Family plays a central role in the lives of adolescents, shaping their choices and well-being. The research places a spotlight on the pivotal role of family support systems in addressing substance abuse and mental health concerns. Adolescents often turn to their families for guidance and solace, making the family unit a vital component in their overall well-being. By fostering these connections, the research suggests that we can empower families to be proactive in addressing substance abuse and mental health concerns within their households.

The research has illuminated the profound impact of peer influence in the lives of adolescents. Peer groups can either exacerbate risky behaviors or serve as a force for positive change. The findings suggest that peer-focused programs hold significant potential to channel this influence constructively. Adolescents often seek validation and acceptance from their peers, making peer groups a powerful avenue for change. By engaging peers as allies in the promotion of mental health and responsible choices, the research suggests that we can tap into a potent resource for change.

The findings of this research represent a substantial contribution to the understanding of adolescent well-being. These insights provide essential guidance for the development of evidence-based policies, intervention strategies, and educational programs, uniquely tailored to adolescents. Policymakers, clinicians, and educators have a richer understanding of the factors at play, equipping them to make informed choices that enhance the lives of adolescents.

In essence, this study is a call to action. It urges us to recognize the specific vulnerabilities of adolescents, to strengthen the support systems provided, and to harness the power of peer influence for the betterment of youth. It challenges us to embrace early interventions and to engage proactively with the complex issues of substance abuse and mental health. By doing so, we can foster a generation of adolescents equipped to face the future with resilience and well-being, ultimately enhancing the overall quality of life for youth.

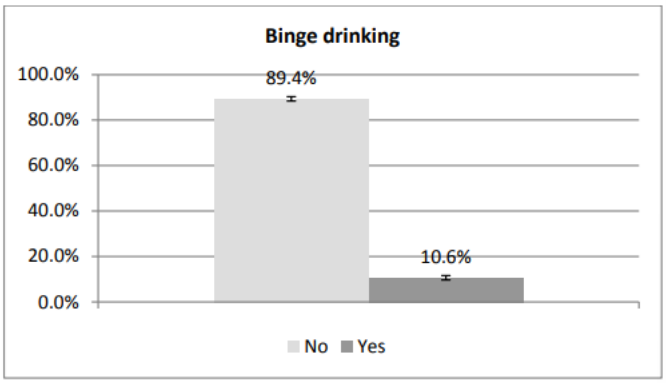
5.0 References:

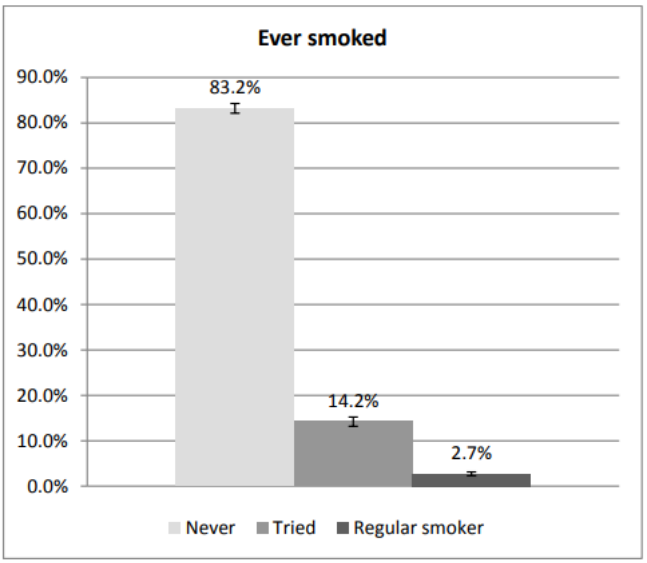
1. Mental Health Foundation. "Children & Young People Statistics." Mental Health Foundation, https://www.mentalhealth.org.uk/explore-mental-health/statistics/children-young-people-statistics. Accessed 27 Aug. 2023.
2. McHugh ML. The chi-square test of independence. Biochem Med (Zagreb). 2013;23(2):143-9. doi: 10.11613/bm.2013.018. PMID: 23894860; PMCID: PMC3900058.
3. Sarker IH (2022) Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Comput Sci, pp. 1–20
4. Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. Expert Systems With Applications, 134, 93-101. https://doi.org/10.1016/j.eswa.2019.05.028
5. Kessler, R. C., Green, J. G., Gruber, M. J., Sampson, N. A., Bromet, E., Cuitan, M., Furukawa, T. A., Gureje, O., Hinkov, H., Hu, Y., Lara, C., Lee, S., Mneimneh, Z., Myer, L., Sagar, R., Viana, M. C., & Zaslavsky, A. M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: Results from the WHO World Mental Health (WMH) survey initiative. International Journal of Methods in Psychiatric Research, 19(Suppl 1), 4-22. https://doi.org/10.1002/mpr.310
6. Bell, S., Britton, A. An exploration of the dynamic longitudinal relationship between mental health and alcohol consumption: a prospective cohort study. BMC Med 12, 91 (2014). https://doi.org/10.1186/1741-7015-12-91
7. Mair, C., Lipperman-Kreda, S., Gruenewald, P. J., Bersamin, M., & Grube, J. W. (2015). Adolescent Drinking Risks Associated with Specific Drinking Contexts. Alcoholism: Clinical and Experimental Research, 39(9), 1705-1711. https://doi.org/10.1111/acer.12806
8. Windle, M. (2003). Alcohol Use Among Adolescents and Young Adults. Alcohol Research & Health, 27(1), 79-85. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6676696/
9. Romac D, Muslić L, Jovičić Burić D, Orban M, Đogaš V, Musić Milanović S. The Relationship between Alcohol Drinking Indicators and Self-Rated Mental Health (SRMH): Standardized European Alcohol Survey (SEAS). Healthcare. 2022; 10(7):1260. https://doi.org/10.3390/healthcare10071260
10. Craig, J. M., Zettler, H. R., Wolff, K. T., & Baglivio, M. T. (2018). Considering the Mediating Effects of Drug and Alcohol Use, Mental Health, and Their Co-Occurrence on the Adverse Childhood Experiences–Recidivism Relationship. Youth Violence and Juvenile Justice. https://doi.org/10.1177/1541204018796910
11. Barbara S. Thomas & Lan Tien Hsiu (1993) The Role of Selected Risk Factors in Predicting Adolescent Drug Use and Its Adverse Consequences, International Journal of the Addictions, 28:14, 1549-1563, DOI: 10.3109/10826089309062199
12. Nawi, A.M., Ismail, R., Ibrahim, F. et al. Risk and protective factors of drug abuse among adolescents: a systematic review. BMC Public Health 21, 2088 (2021). https://doi.org/10.1186/s12889-021-11906-2
13. Lindsey A Hines, Hannah J Jones, Matthew Hickman, Michael Lynskey, Laura D Howe, Stan Zammit, Jon Heron. (2023). Adverse childhood experiences are “strong predictor” for adolescent cannabis use.
14. Viner RM, Kinra S, Nicholls D, Cole T, Kessel A, Christie D, White B, Croker H, Wong ICK, Saxena S. Burden of child and adolescent obesity on health services in England. Arch Dis Child. 2018 Mar;103(3):247-254. doi: 10.1136/archdischild-2017-313009. Epub 2017 Aug 1. PMID: 28765261.
15. Conner KR, Pinquart M, Gamble SA (2009) Meta-analysis of depression and substance use among individuals with alcohol use disorders. J Substance Abuse Treatment 37(2):127–137
16. Barrett AE, Turner RJ (2006) Family structure and substance use problems in adolescence and early adulthood: examining explanations for the relationship. Addiction 101(1):109–120

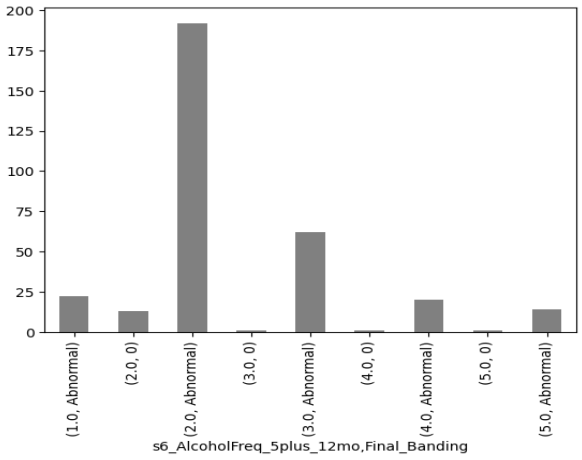
17. Pierce JP, Distefan JM, Kaplan RM, Gilpin EA (2005) The role of curiosity in smoking initiation.Addict Behav 30(4):685–696
18. Fisher Lisa A, Elias Jeffrey W, Ritz Kathy (1998) Predicting relapse to substance abuse as a function of personality dimensions. Alcoholism: Clin Exp Res 22(5):1041–1047
19. Heydarabadi AB, Ramezankhani A, Barekati H, Vejdani M, Shariatinejad K, Panahi R, Kashfi SH, Imanzad M (2015) Prevalence of substance abuse among dormitory students of shahid beheshti university of medical sciences, Tehran, Iran. Int J High Risk Behav Addict, 4(2)
20. Shi Y (2021) Adv Big Data Anal Theory. Algorithms and Practices. Springer Nature, Berlin
21. Patel D, Shah D, Shah M (2020) The intertwine of brain and body: a quantitative analysis on how big data influences the system of sports. Annal Data Sci 7:1–16
22. McHugh ML (2013) The chi-square test of independence. Biochemia Medica: Biochemia Medica 23(2):143–149
23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: a Machine learning in Python. J Mach Learn Res 12:2825–2830
24. Sarker IH (2021) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2(6):1–20

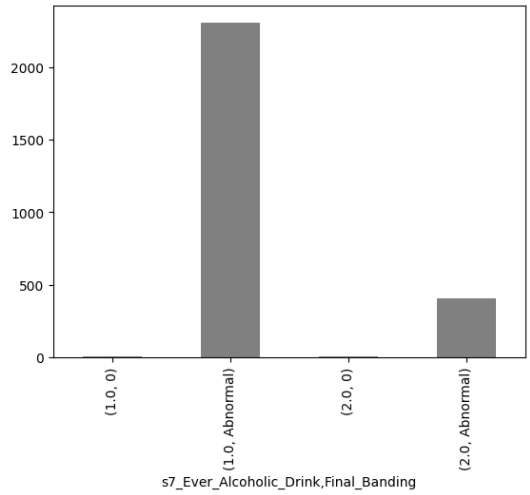
6. APPENDIX

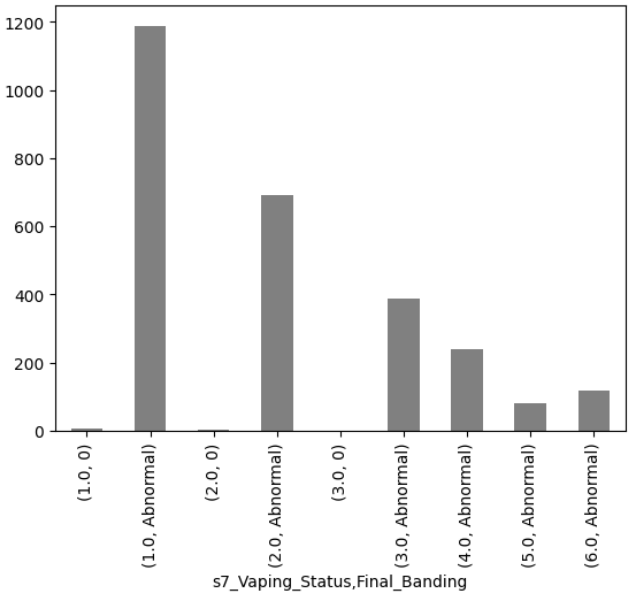
6.1 Bar Graphs

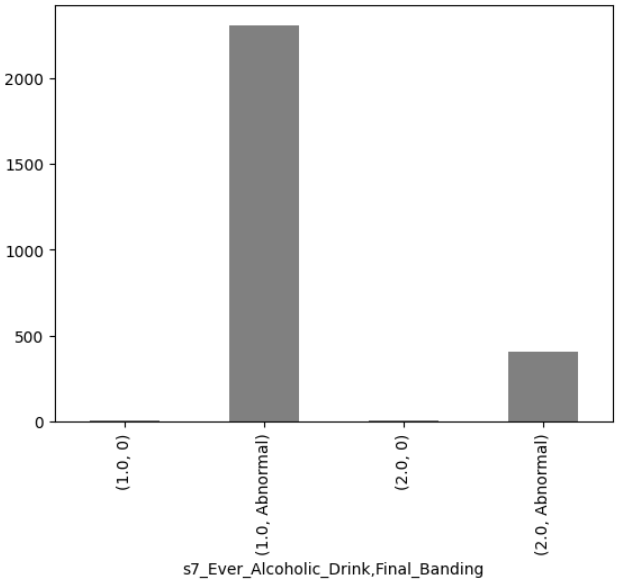












6.2 Comparison Model Results

Logistic Regression Model

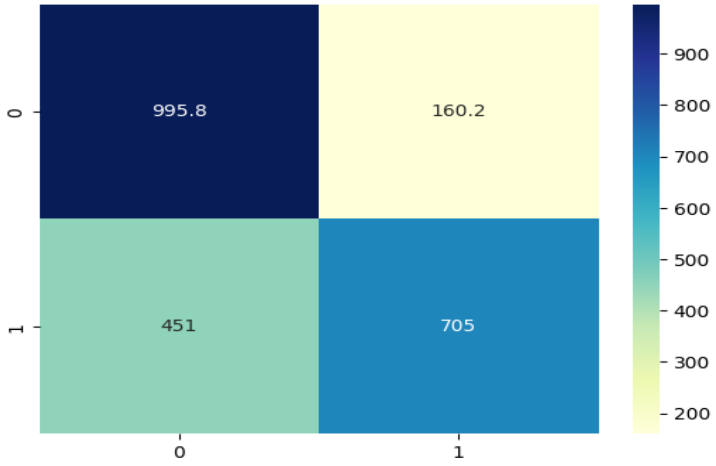
Accuracy: 0.7356401384083044

Precision: 0.8146082623802753

Recall: 0.6097118697317915

F1: 0.6973589152032955

Roc AUC Score: 0.8168703796383165



Naïve Bays Model

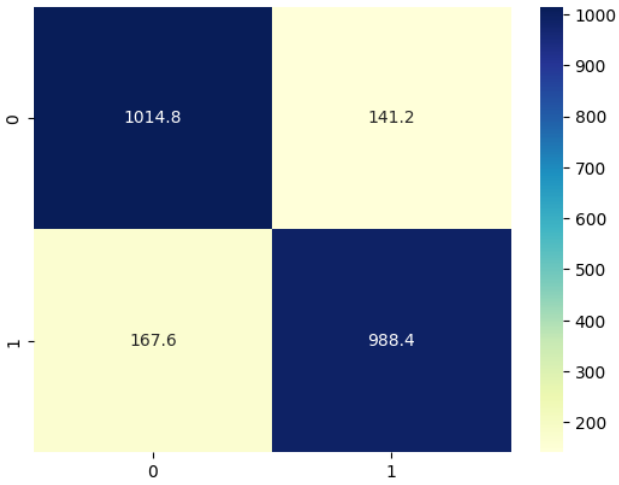
Accuracy: 0.866522491349481

Precision: 0.8764409128639444

Recall: 0.8553349510446218

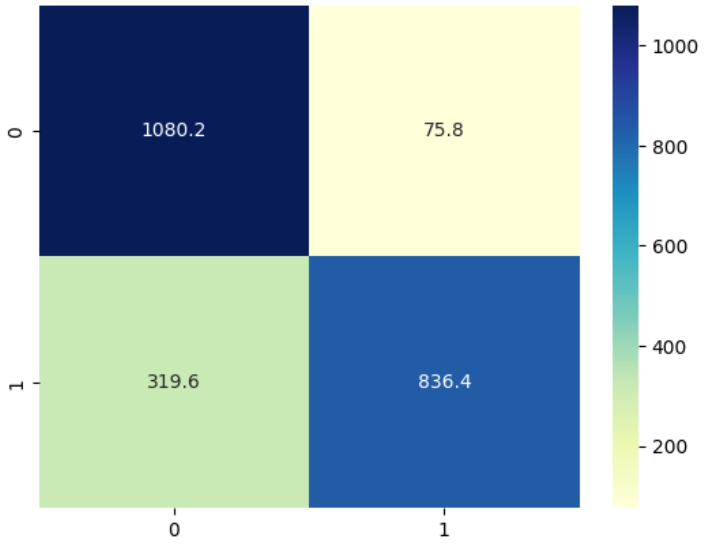
F1: 0.8644662380829503

Roc AUC Score: 0.9047557514123526



Decision Tree Classifier

Accuracy: 0.8289792387543252
Precision: 0.9170721732016026
Recall: 0.7232785764070191
F1: 0.8085353630359567
Roc AUC Score: 0.8587462976028246



Random Forest Model

Accuracy: 0.879325259515571
Precision: 0.8948448135837204
Recall: 0.8574845510272067
F1: 0.8760891662055188
Roc AUC Score: 0.9458450615428592