Project Milestone 2: Analyzing Youth Tobacco Use Trends

Lakshman Kushal Bogi   *Vijeth Jayashekar*   *Meghana Katta*

George Mason University

I. Abstract

Youth tobacco use remains a significant public health concern, necessitating a nuanced understanding of its dynamics to inform effective intervention strategies. This project aims to dissect the complex landscape of youth tobacco use by analyzing the comprehensive Youth Tobacco Survey (YTS) data provided by the Centers for Disease Control and Prevention (CDC). By delving into this rich dataset, the researchers aim to uncover the underlying trends, drivers, and geographical variations in youth tobacco consumption. Given the escalating health risks associated with tobacco use among adolescents, the importance of this research endeavor cannot be overstated. Through meticulous problem delineation and a preliminary literature search, the researchers underscore the urgency of understanding the evolving patterns of tobacco use among youth. Their proposed approach encompasses a multifaceted analytical framework, including data preprocessing, exploratory analysis, statistical modeling, geospatial examination, and insightful interpretation, all aimed at illuminating the complexities of youth tobacco use.

Project Proposal: Analyzing Youth Tobacco Use Trends

II. Introduction

Smoking among young people is a big problem all around the world. It's not good for their health, and it costs a lot for healthcare. Even though people are trying hard to stop it, we still don't fully understand why young people start smoking and keep doing it. This project is about looking very closely at how young people use tobacco. We're using a lot of information collected by the Centers for Disease Control and Prevention (CDC) in something called the Youth Tobacco Survey (YTS).

We want to dig deep into this information to see exactly how many middle and high school students use tobacco and what makes them do it. We're not just looking at the big picture; we want to understand all the small things that affect whether a young person starts smoking or stops. We also want to see if there are differences between places, like cities or states, in how many young people smoke and why.

Our goal is to make sense of all this information so we can come up with better ways to stop young people from smoking. We want to find out what works best to keep them healthy and away from tobacco. By understanding why young people smoke and where they smoke the most, we can help make communities healthier and prevent diseases caused by smoking in the future.

III. Research Question

What are the current trends in youth tobacco use, including prevalence rates, types of tobacco products used, and demographic disparities, based on analysis of the Youth Tobacco Survey (YTS) data?

IV. Problem Statement

The research question seeks to investigate the present state of youth tobacco use, focusing on various aspects such as prevalence rates, types of tobacco products used, and demographic disparities. By analyzing the comprehensive Youth Tobacco Survey (YTS) dataset, this research aims to uncover the latest trends and patterns in tobacco consumption among young people. Understanding these trends is crucial for developing targeted interventions

and policy measures to address the issue effectively. This analysis will provide insights into which tobacco products are most used by youth, how prevalent tobacco use is among different demographic groups, and whether there are disparities in usage rates based on factors such as age, gender, ethnicity, or socioeconomic status.

It is important to conduct this analysis for several reasons. Firstly, youth tobacco use is a significant public health concern with long-term implications for individual health outcomes and broader societal well-being. By gaining a better understanding of the current trends in youth tobacco use, policymakers, public health officials, and other stakeholders can develop more informed strategies and allocate resources effectively to prevent and reduce tobacco consumption among young people. Additionally, identifying demographic disparities in tobacco use can help target interventions towards vulnerable groups who may be at higher risk of tobacco-related harm. Ultimately, this analysis has the potential to contribute to efforts aimed at curbing youth tobacco use and promoting healthier behaviors among young populations.

V. Dataset

The data we used comes from the CDC Youth Tobacco Survey (YTS). This survey collects information about tobacco use among middle and high school students in the United States. It tells us things like how many students use tobacco, what kinds of tobacco they use, and if there are differences between different groups of students. The survey happens every few years, so we can see how things change over time and compare different areas. By looking at this data, we can better understand youth tobacco use and come up with ways to help prevent it.

VI. Problem Description

The prevalence of tobacco use among youth presents a grave concern due to its profound health ramifications, spanning from immediate risks such as addiction and respiratory ailments to long-term vulnerabilities towards chronic diseases in later stages of life. The addictive nature of tobacco compounds coupled with the developing physiology of young individuals exacerbates the risk, potentially leading to lifelong dependencies and enduring health complications. Despite concerted efforts to mitigate tobacco use rates through various interventions and public health campaigns, its prevalence persists as a stubborn challenge, indicating the need for more nuanced and targeted approaches.

Understanding the current landscape of youth tobacco use extends beyond merely quantifying usage rates; it necessitates a comprehensive examination of evolving trends over time and the nuanced variations observed across different demographic groups and geographic regions. By delving into these intricacies, we can uncover patterns of tobacco consumption among youth, identify high-risk populations, and discern the contextual factors driving these behaviors. Such insights are essential for developing effective prevention strategies tailored to address the specific needs and challenges faced by different demographic groups and geographic areas. Only by gaining a holistic understanding of the multifaceted dimensions of youth tobacco use can we devise comprehensive and impactful interventions to curb its prevalence and safeguard the health and well-being of young populations.

VII. Importance of the Problem

The initiation of tobacco use frequently commences during adolescence, rendering young individuals a pivotal demographic for targeted prevention endeavors. Early intervention during this critical developmental stage holds immense significance as it has the potential to disrupt the establishment of ingrained smoking habits, thereby averting the entrenchment of lifelong dependencies and mitigating the myriad health risks associated with tobacco consumption. By intervening early and equipping youth with the

necessary knowledge and resources to resist tobacco initiation, we can effectively shield them from the detrimental health consequences that often accompany long-term tobacco use.

Furthermore, the imperative to reduce youth tobacco use extends beyond individual health concerns and aligns closely with broader public health objectives aimed at fostering healthier communities and alleviating burdens on healthcare systems. By curbing tobacco use among young people, we not only safeguard their individual health but also contribute to the collective well-being of society at large. Reduced tobacco use translates to decreased incidences of tobacco-related diseases, leading to lower healthcare costs and alleviating strains on healthcare infrastructures. Moreover, by promoting healthier behaviors among youth, we lay the foundation for a healthier future generation, fostering resilience against preventable health risks and enhancing overall community well-being. Thus, addressing youth tobacco use emerges as a crucial endeavor with far-reaching implications for both individual health outcomes and societal welfare.

Website Link:

https://mason.gmu.edu/~vjayashe

VIII. Exploratory Data Analysis and Modelling Results

Research Question:

1. How has the distribution of smoking status among youth changed over the years?
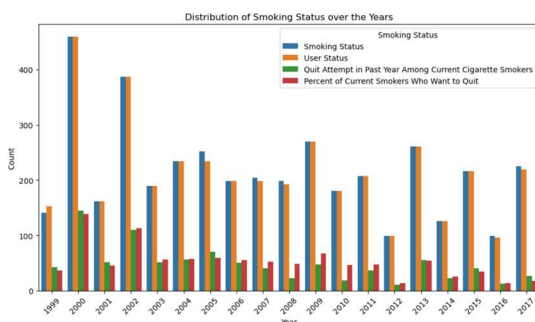

Fig 1. Distribution of Smoking status among youth

2. How does the sample size vary across different smoking statuses among youth?
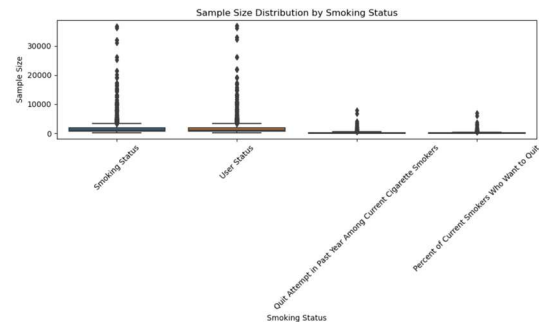

Fig 2. Sample Size Distribution by Smoking Status

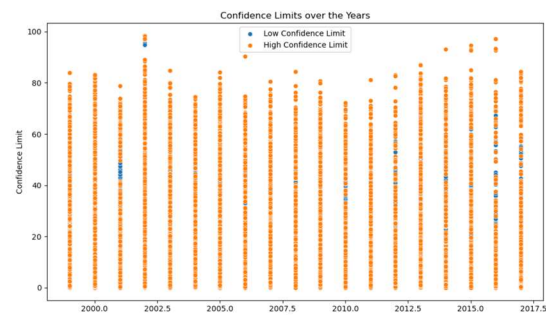**3. How do the confidence limits for youth tobacco use estimates change over the years?**

Fig 3. Confidence Limits over the years

4. What are the relationships between different numerical variables such as year, sample size, data value, low confidence limit, and high confidence limit in youth tobacco use data?
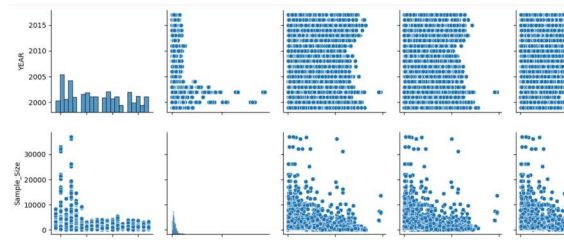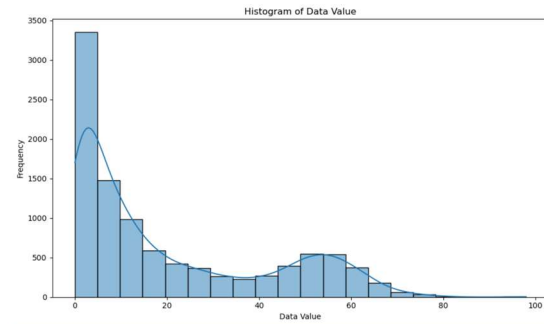
Fig 4. Relationships between numerical variables
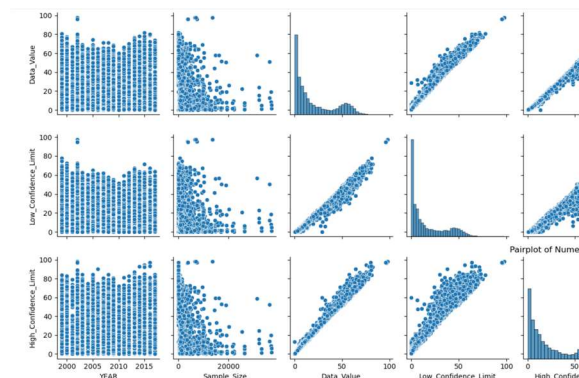

Fig 7. Distribution of Data Values

Fig 5. Relationship between numerical variables

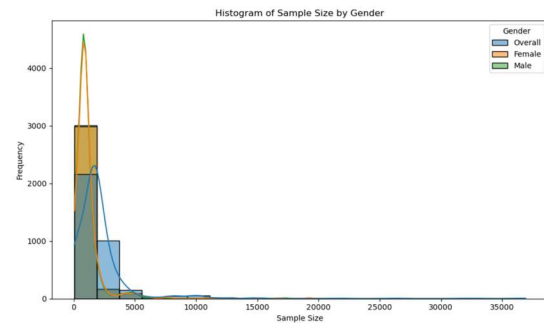**7. How does the sample size vary by gender in youth tobacco use data?**

Fig 8. Sample Size Variation by Gender

5. Is there any relationship between the sample size and the data value in youth tobacco use data?

8. What is the correlation between different numerical variables in the youth tobacco use dataset, and how does this correlation manifest in a heatmap?
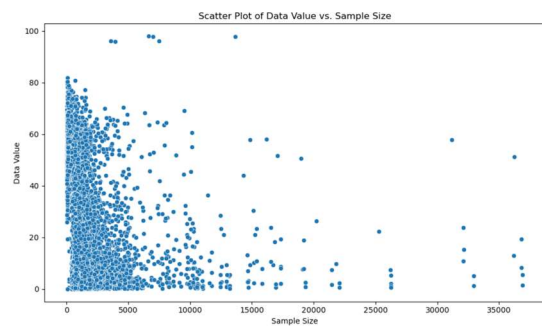

Fig 6. Data Values vs Sample Size

6. What is the distribution of data values in youth tobacco use data?
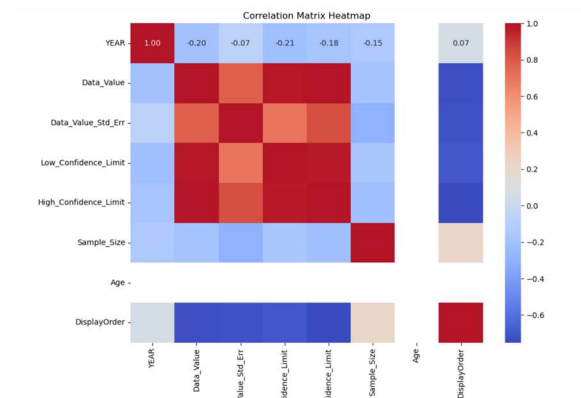
Fig 9. Correlation between Numerical Variables

9. How does the distribution of tobacco use data values vary among different racial groups in the dataset?
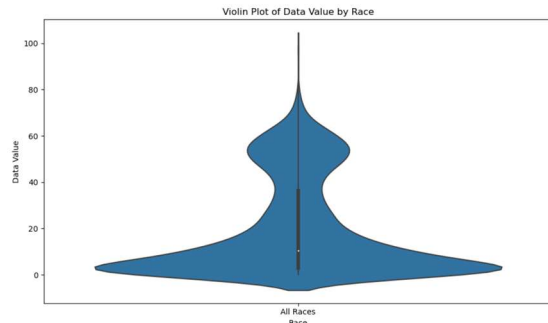
Fig 10. Data Value Variation by Racial Groups

10. How do the relationships between numerical variables (such as Sample_Size, Data_Value, Low_Confidence_Limit, High_Confidence_Limit) differ between genders in the dataset?
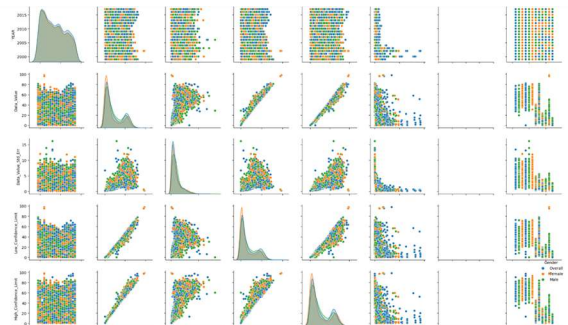
Fig 11. Relationships between numerical variables(difference between genders)
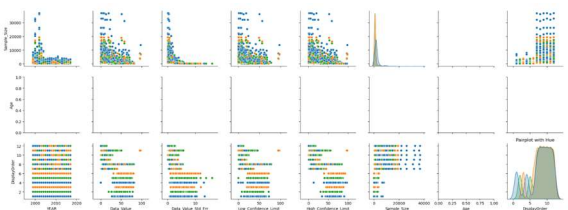
Fig 11. Relationships between numerical variables (difference between genders)

IX. Data Preprocessing:

Figure 13

Fig 13. Head

Fig 14. Missing values

```
            YEAR    Data_Value  Data_Value_Std_Err  Low_Confidence_Limit \
count  10076.000000  10076.000000        10076.000000          10076.000000
mean    2006.867705     20.319035            1.850973             16.703374
std        5.417412     21.153165            1.777582             18.559545
min     1999.000000      0.000000            0.000000              0.000000
25%     2002.000000      3.000000            0.600000              1.700000
50%     2006.000000     10.500000            1.300000              7.900000
75%     2011.000000     36.400000            2.400000             29.600000
max     2017.000000     98.000000           16.100000             97.600000

       High_Confidence_Limit  Sample_Size        Gender     Race      Age \
count           10076.000000  10076.000000  10076.000000  10076.0  10076.0
mean               23.943132   1483.832572      1.017467      0.0      0.0
std                23.967822   2189.891960      0.819909      0.0      0.0
min                 0.000000     50.000000      0.000000      0.0      0.0
25%                 4.300000    671.000000      0.000000      0.0      0.0
50%                13.200000   1001.500000      1.000000      0.0      0.0
75%                42.325000   1642.000000      2.000000      0.0      0.0
max                98.400000  36910.000000      2.000000      0.0      0.0

          Education  DisplayOrder
count  10076.000000  10076.000000
mean       0.511513      8.343092
std        0.499892      2.945307
min        0.000000      1.000000
25%        0.000000      7.000000
50%        1.000000      9.000000
75%        1.000000     11.000000
max        1.000000     12.000000
```

Fig 15: Summary statistics of the dataset
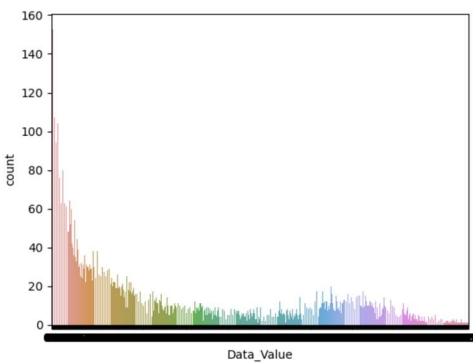
Fig 16. The distribution of the target variable
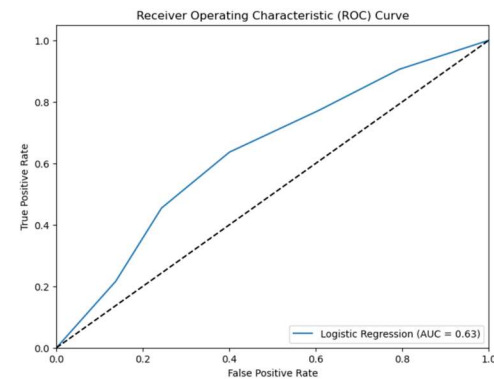
X. Modeling

How does demographic information such as gender, race, age, and education level relate to the likelihood of a youth having a tobacco usage rate exceeding a certain threshold (e.g., 20%)?

XI. Techniques Used

1. *Logistic Regression Model*:
- Technique: Logistic regression modeling was employed to predict the binary outcome variable based on the selected features.
- Results: The logistic regression model achieved an accuracy of approximately 62.85% and an AUC score of 0.63.

Accuracy of Logistic Regression: 0.6284722222222222

Fig 17: Confusion Matric for Logistic Regression

Fig 18: ROC Curve

2. *Decision Tree Model:*
- Technique: Decision tree modeling was utilized to predict the binary outcome variable using a tree-like structure of decisions.
- Results: The decision tree model yielded an accuracy of 65.62%.

Accuracy of Decision Tree: 0.65625

Confusion Matrix for Decision Tree

Fig 19: Confusion Matrix for Decision Tree

Receiver Operating Characteristic (ROC) Curve for Decision Tree
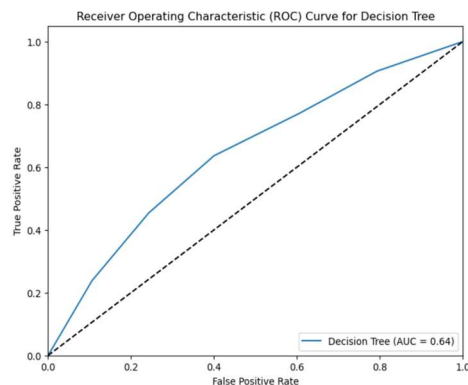
Fig 20. ROC Curve for Decision Tree

XII. Model Improvement for Next Final Paper

For the final research paper, several techniques will be employed to improve the predictive performance of the models:

1. *Hyperparameter Tuning*:
● Utilize techniques like grid search or random search to optimize the hyperparameters of the models, such as regularization parameters for logistic regression and maximum depth for decision trees.
2. *Feature Engineering:*

● Explore various feature engineering techniques, including creating interaction terms, scaling features, and encoding categorical variables more effectively.
3. *Model Cross-Validation:*
● Implement k-fold cross-validation to assess the models' generalization performance and reduce overfitting.
4. *Ensemble Methods:*
● Experiment with ensemble methods like random forests or gradient boosting to combine multiple models for improved predictive accuracy.
5. *Advanced Algorithms:*
● Consider using more advanced machine learning algorithms such as support vector machines or neural networks to capture complex relationships in the data.
6. *Feature Selection:*
● Perform feature selection techniques to identify the most informative features and reduce dimensionality, thus enhancing model interpretability and efficiency.
7. *Model Evaluation:*
● Conduct comprehensive model evaluation, including assessing additional metrics such as precision, recall, and F1-score, to gain a holistic understanding of the models' performance.

By implementing these techniques, the aim is to develop more robust and accurate predictive models that can effectively address the research question or problem statement outlined in the final research paper.

XIII. Preliminary Literature Search

The literature on machine learning applications for predicting smoking cessation outcomes among diverse populations has demonstrated significant advancements in recent years. Studies such as those conducted by Lai et al. (2018), Coughlin et al. (2017), and Medina and Mohaghegh (2020) have employed machine learning algorithms to analyze various datasets and identify predictors of smoking cessation success. These predictors often encompass a wide

range of factors, including demographic characteristics, smoking behavior patterns, socio-economic status, and psychosocial factors. By leveraging large-scale datasets and advanced analytical techniques, researchers aim to develop accurate predictive models that can inform the design and implementation of targeted smoking cessation interventions. Additionally, studies such as Hébert et al. (2021) have focused on predicting smoking relapse events among individuals who have successfully quit smoking, highlighting the importance of personalized interventions to prevent relapse and sustain smoking cessation outcomes. Moreover, Kulikova et al. (2019) explored the potential of leveraging social media data for real-time monitoring and prediction of smoking-related behaviors, underscoring the utility of novel data sources in enhancing predictive modeling efforts. Overall, the literature underscores the critical role of machine learning in advancing our understanding of smoking cessation dynamics and facilitating the development of tailored interventions to address tobacco use and improve public health outcomes. (Issabakhsh et al, 2023)

The paper titled "Data mining classification techniques: an application to tobacco consumption in teenagers" authored by Juan J. Montaño-Moreno, Elena Gervilla-García, Berta Cajal-Blasco, and Alfonso Palmer presents an in-depth analysis of the predictive power of various psychosocial and personality variables on nicotine consumption among adolescents. The study utilizes different classification techniques derived from Data Mining to discern patterns and relationships within a dataset of teenage tobacco consumers and non-consumers. The introduction sets the stage by highlighting the persistent challenge of tobacco addiction, particularly among adolescents, despite advancements in prevention and treatment efforts. It emphasizes the detrimental consequences of tobacco use initiated during adolescence, including increased susceptibility to other substance abuse. The introduction also underscores the complex interplay of familial, peer-related, and personality factors in shaping adolescent tobacco consumption behavior, as evidenced by numerous studies cited. The authors discuss the significance of understanding the multifaceted nature of addictive behaviors and the limitations

of traditional statistical methods in capturing complex relationships within large datasets. They introduce Data Mining as a promising approach, citing its ability to handle vast amounts of data and identify intricate patterns automatically. The literature review delves into the theoretical underpinnings of Data Mining techniques, particularly Artificial Neural Networks (ANNs), decision trees, logistic regression, and discriminant analysis. The authors provide a comprehensive overview of each technique, elucidating their respective strengths and applications in predictive modeling. Drawing on previous studies in the field of addictive behaviors, the authors demonstrate the efficacy of Data Mining techniques in predicting drug consumption and tailoring treatment interventions. They highlight the versatility of ANNs in capturing nonlinear relationships, decision trees' descriptive nature, and logistic regression's suitability for categorical data analysis. Furthermore, the literature review contextualizes the current study within the broader research landscape, noting the scarcity of studies applying Data Mining techniques to adolescent tobacco consumption. It underscores the novelty of the research endeavor and the potential contributions of the study findings to the field of tobacco control and prevention. In summary, the literature review provides a comprehensive synthesis of existing research on adolescent tobacco consumption, traditional statistical methods, and the emerging field of Data Mining. It sets the stage for the current study's exploration of predictive modeling techniques to discern patterns of tobacco consumption behavior among teenagers, thereby contributing to the development of effective prevention strategies (Data Mining Classification Techniques: An application., 2023)

The literature surrounding nicotine addiction among youth, particularly in the context of e-cigarette and hookah use, underscores the escalating prevalence of these behaviors and the imperative for effective intervention strategies. Previous research has extensively examined socio-environmental determinants that contribute to nicotine addiction, revealing factors such as early initiation, peer influence, family smoking behavior, and socio-

economic status as significant predictors. However, traditional approaches to prediction and intervention often fall short in capturing the complexity of these relationships. In recent years, the application of machine learning algorithms in public health research has emerged as a promising avenue for improving predictive modeling and risk assessment in tobacco control and substance use prevention. Studies utilizing machine learning techniques have demonstrated notable advancements in identifying predictive factors for nicotine addiction among youth, offering insights into nuanced patterns and interactions that may inform targeted intervention efforts. Against this backdrop, the current study aims to contribute to the existing literature by identifying predictor variables and developing robust nicotine addiction prediction models tailored specifically to youth e-cigarette and hookah users. By leveraging machine learning methodologies, this research seeks to enhance understanding of the underlying determinants of youth tobacco use and pave the way for more effective prevention and intervention strategies tailored to the unique needs of this population. The study, conducted by Jeeyae Choi, Hee-Tae Jung, Anastasiya Ferrell, Seoyoon Woo, and Linda Haddad, employs machine learning algorithms, specifically Random Forest with ReliefF and Least Absolute Shrinkage and Selection Operator (LASSO), to analyze data from the National Youth Tobacco Survey (Choi et al., 2021)

The prevalence of tobacco use among youth presents a significant public health concern globally, with profound implications for individual health outcomes and societal well-being. Satpathy et al. (Year) examine the gender dimensions of youth vulnerability to accessing cigarettes in South-East Asia (SEA) using data from the Global Youth Tobacco Survey (GYTS). Their study underscores the urgent need to understand the socio-cultural, economic, and legal factors influencing youth tobacco use and access to tobacco products. Despite efforts to curb tobacco initiation, evidence suggests widespread prevalence, with a considerable proportion of smokers initiating tobacco use during adolescence. Gender disparities in tobacco use are evident, with higher prevalence rates

among boys compared to girls, though a positive correlation between smoking prevalence among genders exists. Various methods of tobacco access, including physical, financial, and illegal avenues, underscore gaps in enforcement and compliance with tobacco control policies, despite legal restrictions on sales to minors. The study highlights the importance of comprehensive tobacco control policies, guided by international frameworks such as the WHO Framework Convention on Tobacco Control (FCTC), in mitigating youth tobacco use and access. Article 6 (tobacco taxation), Article 13 (advertising bans), and Article 16 (sales restrictions to minors) of the FCTC play crucial roles in shaping effective tobacco control strategies aimed at protecting youth from tobacco initiation and addiction. Moving forward, further research is warranted to delve deeper into the socio-cultural and economic determinants of youth tobacco use, facilitating evidence-based interventions to address this pressing public health issue. (Satpathyet al., 2022)

The Summary Results of the Global Youth Tobacco Survey (GYTS) in selected countries of the WHO European Region reveal the persistent public health challenge posed by tobacco use among young people aged 13-15. Despite efforts to curb tobacco prevalence, over 8 million deaths annually are attributed to tobacco-related illnesses, with most smokers initiating tobacco use during adolescence. The GYTS, a standardized survey, collects data on tobacco use and control indicators. Findings from 25 countries underscore variations in tobacco prevalence, with higher rates among boys, although prevalence among girls is rising in some regions. Despite declines in tobacco use in some countries over survey periods, others show stagnation or increases. Exposure to secondhand smoke at home is declining, but remains concerning, while access to cigarettes for minors remains high, indicating lax enforcement of age restrictions. Marketing tactics, including advertisements and free product distributions, target youth extensively, underscoring the need for comprehensive tobacco control measures aligned with the WHO Framework Convention on Tobacco Control. Efforts to prevent tobacco initiation among young people are paramount,

necessitating robust policies, smoke-free environments, cessation support, and protection from industry influences. Full implementation of WHO FCTC measures, coupled with multisectoral coordination, is crucial in combatting the tobacco epidemic and safeguarding public health (World Health Organization, 2023).

The systematic review and meta-analysis titled "Association Between Exposure to Tobacco Content on Social Media and Tobacco Use" conducted by Scott I. Donaldson, PhD, MS, Allison Dormanesh, MS, Cindy Perez, BA, Anuja Majmundar, PhD, MBA, and Jon-Patrick Allem, PhD, MA, aimed to investigate the relationship between exposure to tobacco-related content on social media and tobacco use behaviors. The study, published in JAMA Pediatrics, analyzed 29 relevant studies to assess whether exposure to tobacco content on social media platforms influenced lifetime tobacco use, past 30-day tobacco use, and susceptibility to tobacco use among individuals who had never used tobacco before. The authors identified a substantial body of evidence suggesting that exposure to tobacco content on social media was associated with increased odds of reporting lifetime tobacco use, past 30-day tobacco use, and susceptibility to tobacco use among never users. The findings were consistent across various demographic groups, including adolescents and young adults, indicating that exposure to tobacco-related content on social media platforms posed a significant risk factor for initiating and continuing tobacco use behaviors. The study's comprehensive approach included a systematic search across multiple databases and thorough screening of articles based on predefined inclusion and exclusion criteria. The authors employed rigorous statistical analyses, including random-effects meta-analysis and subgroup analyses, to assess the heterogeneity of effect sizes and explore potential sources of variation among studies. The review highlighted the pervasive nature of tobacco-related content on popular social media platforms such as Twitter, Facebook, Instagram, and YouTube, indicating that these platforms served as avenues for tobacco companies to promote their products and influence users' attitudes and behaviors toward

tobacco use. The authors emphasized the need for regulatory action to address the presence of tobacco content on social media and mitigate its impact on adolescent and young adult populations. Overall, the study provided valuable insights into the association between exposure to tobacco content on social media and tobacco use behaviors, underscoring the importance of targeted interventions and policy measures to curb the influence of pro-tobacco content in online environments (Scott I. Donaldson, 2022).

The literature review for the paper "Tobacco use by youth: a surveillance report from the Global Youth Tobacco Survey project" integrates the contributions of the authors: Charles W. Warren, Leanne Riley, Samira Asma, Michael P. Eriksen, Lawrence Green, Curtis Blanton, Cliff Loo, Scott Batchelor, and Derek Yach. These researchers have collectively illuminated the global landscape of youth tobacco consumption, leveraging the findings of the Global Youth Tobacco Survey (GYTS) to inform public health policies and interventions aimed at curbing adolescent smoking rates. It begins by highlighting the alarming global burden of tobacco-related mortality, emphasizing its status as a leading cause of preventable death. The review underscores the World Health Organization's estimation of approximately 4 million tobacco-related deaths annually, a figure projected to escalate to 8.4 million by 2020, with 70% of these fatalities anticipated to occur in developing nations. Moreover, it elucidates the early onset of tobacco use, with a substantial portion of individuals initiating smoking before reaching 18 years of age. Recent trends indicate an upward trajectory in smoking prevalence among children and teenagers, indicating the urgency of addressing this public health issue. The introduction of the Global Youth Tobacco Survey (GYTS) initiative is outlined, emphasizing its collaborative development by the World Health Organization (WHO) and the US Centers for Disease Control and Prevention (CDC). The GYTS aims to assess youth tobacco consumption across various countries using a standardized methodology and questionnaire. Employing a two-stage sampling process within schools, the survey gathers representative data on smoking habits among 13 to 15-year-old

students. The literature review further outlines the survey's implementation in twelve countries, including Barbados, China, Costa Rica, Russia, Fiji, Jordan, Poland, Sri Lanka, South Africa, Ukraine, Venezuela, and Zimbabwe. Key findings from these surveys highlight smoking prevalence rates ranging from 10% to 33% among youth, with a majority expressing a desire to quit smoking despite limited access to cessation programs. Additionally, the review underscores the pervasive influence of tobacco advertising in the media and the significant exposure to environmental tobacco smoke among adolescents. The literature review underscores the critical role of the GYTS in providing valuable insights into youth tobacco consumption patterns, prompting other countries to seek assistance in conducting similar surveys. It highlights ongoing collaborative efforts by WHO, CDC, and other organizations to support global tobacco control initiatives. Ultimately, the review underscores the importance of standardized surveillance systems like GYTS in evaluating the effectiveness of national tobacco control programs and the WHO Framework Convention on Tobacco Control, particularly concerning safeguarding the health of children and adolescents (scielosp, 2021).

The literature surrounding machine learning applications for predicting smoking addiction among youths has witnessed significant development in recent years. Research such as that conducted by Shreerudra Pratik et al. (2022) has utilized advanced machine learning techniques, including Elastic Net regression and K-Nearest Neighbor (KNN) classification, to identify demographic factors and behavioral patterns associated with smoking addiction among young individuals. This study addresses the pressing concern of the increasing prevalence of smoking addiction among youths, particularly due to the availability of various nicotine products such as E-cigarettes, conventional cigarettes, hookahs, and others. Previous studies have highlighted the growing popularity of E-cigarettes among young adults, with concerns regarding their potential health risks and addictive properties (Pratik et al., 2022). Despite efforts to raise awareness about the dangers of tobacco use, the allure of these products remains high among youths. This underscores the need for predictive models that can effectively identify risk factors and facilitate targeted intervention strategies to curb smoking addiction. Machine learning techniques offer a promising approach to analyzing large-scale survey data, such as the National Youth Tobacco Survey (NYTS), to extract meaningful insights into smoking behavior among youths (Pratik et al., 2022). By leveraging Elastic Net regression for feature selection and KNN for classification, researchers can accurately predict smoking addiction based on various socio-demographic factors, usage patterns, and attitudes toward tobacco products. The predictive models developed in this study demonstrate high performance in terms of accuracy, precision, and recall, indicating their potential utility in identifying individuals at risk of smoking addiction (Pratik et al., 2022). Furthermore, the selected features reveal important insights into the factors influencing smoking behavior among youths, including peer influence, curiosity about tobacco products, exposure to advertisements, and language spoken at home. Despite the progress made in predictive modeling for smoking addiction, challenges remain, including data preprocessing complexities and the need for more robust feature selection techniques (Pratik et al., 2022). Future research efforts may focus on refining preprocessing pipelines for survey data like NYTS and exploring alternative machine learning algorithms to enhance prediction accuracy. Overall, the literature underscores the importance of leveraging machine learning approaches to address the complex dynamics of smoking addiction among youths. By identifying predictive factors and developing targeted intervention strategies, researchers and public health practitioners can work towards mitigating the adverse effects of tobacco use and promoting healthier lifestyles among young populations (Tara Mantler, 2020).

The paper "Comparison and Analysis of Youth Tobacco Surveillance Systems: Lessons Learned and Future Implications" authored by Ichhya Pant provides a comprehensive review of existing literature on youth tobacco surveillance systems. Pant begins by emphasizing the critical role of surveillance in understanding and curbing

tobacco addiction among adolescents. Drawing on the work of public health agencies such as the World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC), the paper highlights the alarming prevalence of tobacco use among youth globally and the associated health risks. Pant delves into the initiation of tobacco use among adolescents, emphasizing the behavioral establishment of smoking habits before the age of 18. By citing studies conducted by CDC and WHO, Pant elucidates the pathways to nicotine addiction and the challenges posed by early smoking initiation. The review underscores the need for accurate and up-to-date surveillance systems to monitor tobacco consumption patterns among youth, emphasizing the significance of data-driven policies and interventions. Furthermore, Pant explores the factors associated with adolescent smoking, including socioeconomic status, peer influence, exposure to smoking in media, and parental support. By synthesizing findings from various studies, Pant underscores the multifaceted nature of adolescent tobacco use and its detrimental impact on public health and economic development. The literature review also discusses existing tobacco surveillance systems, with a focus on the Global Youth Tobacco Survey (GYTS) and the Global School-based Health Survey (GSHS) initiated by WHO and CDC. Pant evaluates the methodologies, survey designs, and dissemination of these surveillance systems, shedding light on discrepancies within and between datasets. Through a meticulous analysis of prior research, Pant identifies gaps and challenges in youth tobacco surveillance and emphasizes the importance of standardization and quality control measures. The review sets the stage for Pant's own research, which aims to compare and analyze different versions of GYTS and GSHS to derive lessons and implications for future tobacco surveillance efforts. In summary, Pant's literature review provides a comprehensive overview of the landscape of youth tobacco surveillance, highlighting the urgency of addressing the tobacco epidemic among adolescents and the need for robust surveillance systems to inform effective public health interventions (Ichhya Pant, 2023).

The paper titled "An Application of Fuzzy Clustering on Prevalence of Youth Tobacco Survey" by Hazel Kavili and Gülhayat Gölbaşi Şimşek presents a novel approach to analyzing data from the National Youth Tobacco Survey (NYTS) 2013 using fuzzy clustering algorithms. Tobacco consumption among youth is a significant public health concern, with most tobacco users initiating their habit during adolescence or young adulthood. The authors aim to understand young people's attitudes toward tobacco products by clustering them into groups based on their usage, tendencies, and intentions. The introduction of the paper underscores the importance of data mining and analysis in various fields, emphasizing the significance of grouping or classifying measurements based on their similarity. Clustering algorithms play a vital role in this process, with various approaches available for different types of problems. The authors highlight the versatility of clustering algorithms in fields such as image analysis, machine learning, bioinformatics, and text analysis. The paper provides a comprehensive overview of clustering algorithms, including hierarchical clustering and K-means clustering, elucidating their methodologies and applications. It discusses the concept of fuzzy logic, introduced by Lutfi Asker Zadeh in 1965, which allows for membership values between 0 and 1, enabling a more nuanced representation of data relationships compared to Boolean logic. Focusing on fuzzy clustering, the authors describe Fuzzy C-Means (FCM) as the most popular algorithm for analyzing complex and multi-dimensional datasets with fuzzy or partial relations. FCM assigns partial membership values to data objects with respect to cluster centers, enabling objects to belong to multiple clusters simultaneously. The paper elucidates the mathematical principles behind fuzzy clustering and its application in analyzing the NYTS 2013 data. The methodology section outlines the data and variables used in the study, including Likert scale questionnaire responses from 267 randomly selected participants. Seven questions related to tobacco use and attitudes were examined using fuzzy clustering algorithms implemented in R Studio. The results section presents descriptive statistics of the sample and discusses the fuzzy clustering outcomes. The findings indicate that while most

young people surveyed are not interested in smoking any kind of tobacco product, a small percentage exhibit a tendency to smoke in the future. The application of fuzzy clustering revealed distinct clusters within the dataset, with differences in membership values reflecting varying attitudes and behaviors toward tobacco use among participants. In conclusion, the paper highlights the potential of fuzzy clustering algorithms in uncovering hidden patterns and insights within complex datasets, such as the NYTS 2013 questionnaire responses. By identifying distinct groups based on tobacco-related attitudes and behaviors, the study contributes to our understanding of youth tobacco consumption patterns and informs targeted interventions to address this public health issue (Elsevier, 2016).

The National Institute on Drug Abuse (NIDA) "Monitoring the Future" (MTF) survey is an essential tool for studying drug usage trends and patterns among youth and high school populations in the US. Since its inception in 1975, the study has been carried out yearly and has yielded extensive data on a variety of substances, including alcohol, tobacco, marijuana, and illegal narcotics. This data provides important insights into the changing face of youth substance use.

The poll is one of the most comprehensive and trustworthy sources of data on trends in teenage drug use since it uses a nationally representative sample of eighth, tenth, and twelfth-grade students. Survey questions include participants' usage of different substances, frequency of use, attitudes about perceived dangers, and perceived risks.

The MTF study provides information on trends in youth tobacco use, including prevalence rates of using e-cigarettes, traditional cigarettes, and other tobacco products. It monitors shifts in the commencement of smoking, frequency of use, and attitudes about tobacco use throughout time, giving policymakers, public health experts, educators, and academics useful information.

Important results from the MTF study provide information on current patterns in teenage tobacco use, such as the fact that among teenagers in some age groups, e-cigarette use is more common than traditional cigarette smoking. Furthermore, the survey provides information on individuals at higher risk for tobacco start and dependency by highlighting differences in tobacco use depending on demographic variables, socioeconomic position, and geographic location.

In addition to providing information for public health policy and prevention efforts, the MTF survey data advances our knowledge of the complex variables impacting the tobacco use habits of young people. The MTF survey is essential in addressing the public health issues connected to youth tobacco use and informing evidence-based treatments aimed at reducing the harm caused by tobacco use among children because it detects new trends and tracks changes in tobacco use patterns over time.

All things considered, the MTF survey is a priceless tool for academics and decision-makers who want to understand and deal with teenage tobacco use in the larger framework of drug abuse prevention and public health promotion. Its extensive data and longitudinal nature make it a vital resource for tracking patterns, figuring out risk factors, and creating successful plans to prevent teenage tobacco usage and the negative health effects that come with it(WHO, 2008).

The U.S. Department of Health and Human Services (HHS) released a historic report titled "The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General" that offers a thorough summary of the negative consequences of tobacco usage on one's health. This study, which was released in 2019, honors fifty years of work to prevent tobacco-related harm and provides a thorough analysis of the scientific data regarding smoking's effects.

The Office on Smoking and Health, a section of the National Center for Chronic Disease Prevention and Health Promotion within the Centers for Disease Control and Prevention, is the author of the study. To clarify the health effects, it makes use of a wealth of research that has been done over the previous 50 years, including experimental, clinical, and epidemiological investigations.

The report's main sections examine the physiological, pathological, and behavioral aspects of tobacco use, emphasizing the harm that smoking causes to one's heart, lungs, risk of cancer, chances of conception, and mental and physical health. It provides a thorough grasp of the complex nature of tobacco-related harm by synthesizing data on the harmful qualities of tobacco smoke, the mechanisms underlying nicotine addiction, and the synergistic consequences of co-occurring medical disorders.

The study looks at the consequences of smoking on an individual's health as well as the larger social effects of tobacco use, such as economic costs, differences in the burden of tobacco-related diseases, and the effectiveness of tobacco control policies and programs. It emphasizes how critical it is to implement all-encompassing tobacco control programs that include both prevention and cessation initiatives.

The report's focus on the advancements made in tobacco control over the last 50 years, including the adoption of evidence-based policies like public education campaigns, smoke-free laws, and tobacco taxes, is one of its main contributions. It also draws attention to recurring issues and areas that require more effort, such as lowering the frequency of smoking in communities that are already at risk, addressing inequalities in the health consequences of tobacco use, and fending off new dangers such underage e-cigarette usage.

All things considered, "The Health Consequences of Smoking—50 Years of Progress" is an invaluable tool for campaigners,

researchers, public health specialists, and lawmakers addressing the tobacco use-related public health dilemma. It is a vital tool for furthering efforts to lower the burden of tobacco-related illnesses and deaths because of its authoritative examination of the scientific data, thorough analysis of the health effects of tobacco use, and insights into practical tobacco control initiatives. (Bonnie RJ et al., 2015)

The 2019 study "E-cigarette devices used by high-school youth" by Krishnan-Sarin et al., which was published in the journal Drug and Alcohol Dependence, offers important new information about the prevalence and usage trends of e-cigarettes among high school students. The writers look at the different kinds of e-cigarettes that young people typically use, illuminating the wide variety of goods on the market and any possible health risks.

The study investigates several facets of e-cigarette device use, including device kinds, brands, and features favored by teenagers, using information from surveys given to high school students. To obtain a thorough knowledge of young people's vaping behaviors, the researchers use a mixed-methods approach, integrating quantitative analysis of survey data with qualitative observations.

The study's key findings show that high school students use a variety of e-cigarette devices, from disposable ones to more sophisticated modular ones with configurable capabilities. The authors note the significance of product design and marketing techniques in forming teenage vaping preferences by identifying elements impacting device selection, such as device aesthetics, simplicity of use, and perceived social acceptability.

The report also highlights the e-cigarette market's explosive growth, which is marked by ongoing product diversification and innovation, such as pod-based systems and electronic

nicotine delivery systems (ENDS) with high nicotine concentrations. These trends raise concerns about the possibility of increased young smoking and addiction, as well as the introduction of new health dangers linked to new device technologies.

The study advances our knowledge of the complex relationships behind teenage vaping behaviors and provides guidance for health campaigns aimed at addressing the youth vaping epidemic by clarifying the landscape of e-cigarette devices utilized by high school students. The results highlight the necessity of thorough regulation of e-cigarette products, including limitations on flavorings, marketing strategies, and nicotine concentration, to shield young people from the negative effects of nicotine addiction and encourage better lifestyle choices.

Overall, the study by Krishnan-Sarin et al. sheds light on the frequency and usage patterns of e-cigarettes among high school students, emphasizing the need for continued investigation and surveillance to track and lessen the negative effects of youth vaping on public health. (Krishnan-Sarin et al., 2019)

The US Department of Health and Human Services (HHS) released a groundbreaking report titled "The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General" that offers a thorough summary of the negative consequences of tobacco usage on one's health. The 2019 study, which was released to commemorate 50 years of efforts to address the smoking-related public health crisis, provides a comprehensive analysis of the scientific data on smoking-related diseases and mortality.

The report, which was written by the Office on Smoking and Health (OSH), a branch of the Centers for Disease Control and Prevention (CDC) inside the National Center for Chronic Disease Prevention and Health Promotion,

explains the physiological, pathological, and behavioral effects of tobacco use by drawing on a wealth of research done over the past 50 years. It documents the numerous negative health impacts of smoking on different organ systems and populations by synthesizing data from clinical trials, epidemiological studies, and experimental research.

Important sections of the paper provide a thorough knowledge of the complex nature of tobacco-related harm by examining the cancerous qualities of tobacco smoke, the mechanisms driving nicotine addiction, and the synergistic consequences of co-occurring health disorders. It also looks at the financial toll that smoking takes, the differences in the burden of tobacco-related diseases, and the efficiency of initiatives and policies aimed at reducing tobacco use.

The report's focus on the advancements made in tobacco control over the last 50 years, including the adoption of evidence-based policies like public education campaigns, smoke-free legislation, and tobacco taxes, is one of its main contributions. But it also draws attention to persistent issues and areas that still require attention, such lowering the frequency of smoking in communities that are already at risk, addressing inequalities in the health consequences of tobacco use, and fending off new threats.

All things considered, "The Health Consequences of Smoking—50 Years of Progress" is an invaluable tool for campaigners, researchers, public health specialists, and lawmakers tackling the tobacco use-related public health dilemma. It is a vital tool for furthering efforts to lower the burden of tobacco-related disease and mortality because of its authoritative examination of the scientific data, thorough analysis of the health effects of tobacco use, and insights into practical tobacco control initiatives. (NIDA, 2023)

Considering the ongoing COVID-19 pandemic, the National Institute on Drug Abuse

(NIDA) news release "Reported drug use among teenagers continued to hold below pre-pandemic levels in 2023" notes significant trends in adolescent drug use. Key findings from national surveys and surveillance systems are included in the news release, offering insights into how the epidemic has affected young people's substance use patterns.

The news release highlights several important facts, one of which is that, as of 2023, recorded drug use among teenagers has not increased above pre-pandemic levels. This implies that there hasn't been a discernible rise in teenage drug use relative to pre-epidemic rates, despite the interruptions and difficulties caused by the COVID-19 pandemic. This result is especially positive because it shows that teenagers are resilient and have developed flexible coping strategies during this time of extraordinary global crises.

The news release might also underscore how crucial it is to keep up preventative efforts and provide support services for teenagers, especially those who are dealing with more stress and difficulties because of the pandemic. It might highlight the necessity of focused treatments to treat drug use disorders and encourage healthy coping mechanisms in young people, as well as the significance of continuing study to track and comprehend the pandemic's long-term effects on the substance use behaviors of young people.

Considering the COVID-19 epidemic, the news release offers insightful information about the current situation of adolescent drug use, showing both encouraging trends and persistent issues. By making this information available, NIDA hopes to promote activities that advance the health and well-being of people, direct policy decisions, and enhance public health initiatives. (NIDA, 2023)

Additionally, the Surgeon General's Reports on Smoking and Tobacco Use offer priceless insights into how tobacco control strategies and policies are changing over time.

These publications provide a roadmap for policymakers to develop comprehensive strategies targeted at reducing tobacco use and its related harms by methodically analyzing and synthesizing scientific information. The studies enable governments to implement laws that can lessen the negative effects of tobacco use on public health by outlining sensible policy measures including tobacco taxes, advertising bans, and smoke-free workplaces. Furthermore, they support the inclusion of tobacco control campaigns in more comprehensive public health programs, acknowledging the relationship between tobacco use and other health inequalities as well as social determinants of health.

The Surgeon General's Reports on Smoking and Tobacco Use highlight the significance of ongoing research and surveillance in addressing the tobacco epidemic in addition to their emphasis on prevention and policy. These studies accelerate progress in our knowledge of the mechanisms underlying tobacco addiction, the effectiveness of quitting strategies, and the effects of newly developed tobacco products on public health by encouraging a culture of inquiry and creativity. They offer a pulse check on tobacco usage trends through continuous surveillance operations, facilitating prompt reactions to new opportunities and difficulties. They also support more funding for research infrastructure and teamwork to advance the field and advance tobacco control.

Finally, the Surgeon General's Reports are effective means of advocacy that raise public awareness and inspire group action to stop the tobacco pandemic. These papers connect with a wide range of stakeholders, including lawmakers, healthcare professionals, educators, and community advocates, by reducing complicated scientific findings into understandable language and captivating stories. They generate a wave of support for evidence-based tobacco control programs by serving as an inspiration for local campaigns and collaboration activities. By being widely circulated and increased, these findings provoke thought, challenge inaction, and provide the momentum for a future empty of the terrible

cost of tobacco-related illness and mortality. (Luther L et al., 1964)

The feasibility and potential efficacy of comprehensive tobacco control programs in promoting youth quit smoking are examined in the 2019 Preventive Medicine Reports study "Feasibility and potential effectiveness of comprehensive tobacco control programs in improving youth tobacco cessation" by Henriksen and Schleicher. By concentrating on the effects of multi-component tobacco control measures on youth cessation outcomes, this study fills a significant vacuum in the literature.

The authors evaluate the viability of putting in place comprehensive tobacco control programs and their potential efficacy in helping young people quit smoking by using a mixed-methods approach. The study provides a comprehensive view of the complex dynamics impacting youth tobacco use habits and quitting efforts by utilizing both quantitative data analysis and qualitative observations.

Comprehensive tobacco control initiatives may have a positive effect on teenage tobacco quitting outcomes, according to the study's key results. Through the implementation of several interventions, including regulatory reforms, school-based instruction, community outreach, and cessation counseling, these programs can establish a supportive environment that enables young people to resist the onset of tobacco use and quit smoking.

Furthermore, to effectively support teenagers quitting, the study emphasizes the significance of addressing a variety of tobacco use determinants, such as individual risk factors, contextual effects, and societal norms. Comprehensive programs can maximize their influence on youth tobacco quitting rates by taking a holistic approach to tobacco control that includes both individual-level interventions and more significant policy reforms.

The authors also address the viability of putting such initiatives into action, considering elements like organizational capacity, stakeholder participation, and resource availability. They underline that to guarantee the effective execution and long-term viability of comprehensive tobacco control programs, cooperation between governmental entities, neighborhood associations, educational institutions, and healthcare professionals is essential.

All things considered, Henriksen and Schleicher's research offers insightful information about the viability and possible efficacy of comprehensive tobacco control initiatives in promoting teenage tobacco cessation. The study provides insight into the essential elements and workings of these programs, as well as the opportunities and problems of implementing them in the real world. This helps to build evidence-based solutions that will lower juvenile tobacco use and encourage healthier habits. (Henriksen et al., 2018)

"Impact of social networks and norms on e-cigarette use among teenagers in Southern California" is the title of a prospective cohort study that investigates how social networks and norms affect e-cigarette usage among teenagers in the Southern California area. By investigating how peer networks and social norms influence the onset and progression of e-cigarette use among youth populations, the study seeks to fill in gaps in the literature.

The authors track a group of teenagers using a longitudinal approach to evaluate how e-cigarette use behavior varies over time and how it relates to social factors like peer pressure, societal norms around e-cigarette use, and social network features. Through monitoring personal information about e-cigarette consumption and social networks, the research offers valuable perspectives on the mechanisms of peer influence.

Important conclusions from the research provide insight into how social networks and norms influence teenagers' paths toward e-cigarette use. Peer group attributes, like the frequency of e-cigarette use in social circles and their opinion of vaping's social acceptability, are recognized by the authors as important indicators of an individual's e-cigarette use behavior. They also investigate how teenagers' vulnerability to e-cigarette experimentation and progression to regular use is influenced by changes in social networks across time, including changes in peer connections and social norms.

The study also looks at social learning processes, peer pressure, and social reinforcement as possible mechanisms underpinning how social networks and norms affect teenage e-cigarette use. The authors' explanation of the mechanisms by which social variables influence vaping behaviors through their explanation of how social factors influence vaping habits, the authors offer important insights on how to design young e-cigarette use prevention and intervention programs.

The study's prospective group design makes it possible to identify the temporal links between social factors and the effects of e-cigarette usage, which improves our comprehension of the causal pathways connecting social surroundings to the vaping behaviors of teenagers. The study reduces possible biases associated with cross-sectional research designs by prospectively tracking individuals over time. This approach further strengthens the data supporting the influence of social networks and norms on e-cigarette usage among adolescents.

All things considered, the study offers insightful information about the intricate relationship between social variables and teenage e-cigarette use, emphasizing the significance of addressing social influences in all-encompassing tobacco control initiatives meant to lower the incidence of youth vaping and lessen the negative effects of e-cigarette use on public health. (Piombo SE et al., 2020)

Investigating the effects of extended exposure to vaporized nicotine on affective and cognitive behaviors in male mice, the study "Effect of chronic vapor nicotine exposure on affective and cognitive behavior in male mice" was conducted. Using a preclinical animal model, this study attempts to fill in knowledge gaps about the behavioral effects of long-term nicotine exposure, specifically about mood and cognition.

The authors investigate the effects of extended nicotine inhalation on a variety of behavioral outcomes related to affective and cognitive function using a mouse model of chronic vapor nicotine exposure. The study uses a battery of behavioral tests to evaluate cognitive performance in learning and memory activities, as well as mood-related behaviors like anxiety-like behavior and depressive-like symptoms.

The study's main conclusions provide light on the behavioral effects of long-term nicotine consumption in male mice. After extended exposure to vaporized nicotine, the authors note changes in affective behavior, such as an increase in anxiety-like behavior and depressive-like symptoms. These results indicate that long-term nicotine exposure may negatively impact male mice's ability to regulate their moods and emotional health.

The study also looks at how long-term nicotine exposure affects cognitive function; it finds that mice exposed to nicotine perform worse in learning and memory tasks than mice in control groups. These mental disorders imply that long-term nicotine inhalation may affect male mice's ability to learn, create memories, and make decisions.

The authors also investigate possible underlying mechanisms, such as changes in neurotransmitter systems, neuroinflammatory processes, and neuroplasticity in brain regions related to mood regulation and cognitive function, that may link long-term nicotine exposure to affective and cognitive dysfunction. The work offers important insights into the

pathophysiology of nicotine addiction and its effects on brain function by clarifying the neurobiological mechanisms by which nicotine influences behavior.

Furthermore, controlled testing and variable manipulation are made possible using preclinical animal models, which help researchers determine the causal links between long-term nicotine exposure and behavioral consequences. The study improves our knowledge of the behavioral effects of long-term nicotine exposure and guides efforts to create focused therapies for nicotine addiction and related comorbidities by utilizing rigorous experimental techniques and thorough behavioral assessments.

All things considered, the study adds to the expanding corpus of research on the neurobehavioral effects of nicotine exposure and emphasizes how critical it is to examine both the affective and cognitive domains in studies on nicotine addiction and the harm associated with tobacco use. The study sheds light on how long-term nicotine exposure affects male mice's mood and cognitive function. This helps to understand the behavioral effects of nicotine addiction and guides efforts to reduce the negative health effects of tobacco use.(Murdaugh LB et al,. 2024)

The study "Susceptibility to e-cigarette use and associated factors in high school youth, Oklahoma Youth Tobacco Survey, 2021-2022" investigates the factors that are connected to Oklahoma high school students' susceptibility to using e-cigarettes. This study attempts to shed light on the factors that influence young people's decision to start using e-cigarettes, with an emphasis on the individual, social, and environmental aspects that affect vaping susceptibility.

The authors investigate the incidence of e-cigarette susceptibility among high school students and identify variables linked to increased susceptibility to e-cigarette use using data from the Oklahoma Youth Tobacco Survey, which was carried out in 2021–2022. Using a cross-sectional survey design, the study gathers information on demographics, tobacco usage patterns, peer influences, and other pertinent aspects in addition to assessing students' vulnerability to starting e-cigarettes.

The study's main conclusions show that Oklahoman high school students are concerningly susceptible to using e-cigarettes, and a sizable percentage of them have expressed interest in or plans to attempt vaping in the future. The authors list several variables that are linked to a higher risk of using e-cigarettes, such as exposure to e-cigarette advertising, peer pressure, the perception of social norms surrounding vaping, and personal risk factors like sensation-seeking inclinations or ignorance of the dangers of vaping.

In addition, the study looks at social modeling, peer pressure, and the normalization of vaping practices in social networks as possible factors underlying vulnerability to e-cigarette usage. The authors offer important insights into the complex nature of juvenile e-cigarette initiation by clarifying the social and environmental implications on adolescent vaping susceptibility. They also emphasize the significance of targeting these aspects in preventative efforts.

The results of this study have significant implications for preventative initiatives and tobacco control laws intended to lower the incidence of youth vaping and lessen the negative effects of e-cigarette usage on public health. To prevent high school kids from starting e-cigarettes and to encourage healthy habits, the study provides information for targeted interventions by identifying modifiable risk factors and protective factors related with e-cigarette susceptibility.

Overall, the study advances our knowledge of the factors that influence high school students' susceptibility to e-cigarettes by offering insightful information on the intricate

interactions between social, environmental, and personal factors that affect teenage vaping practices. The study provides evidence-based recommendations for tobacco control and public health promotion efforts targeted at addressing Oklahoma's youth vaping pandemic by identifying critical characteristics linked to vulnerability to e-cigarette usage. (James SA et al., 2024)

XIV. Proposed Approach

The proposed approach involves several key steps:

1) *Data Preparation:* Before we can start looking at the data, we need to get it ready. This means we must clean it up and organize it so that it's easy to understand and trustworthy. We fix any mistakes or errors in the data and make sure all the information is consistent. This step is important because if the data isn't clean and reliable, our analysis won't be accurate.

2) *Exploratory Data Analysis:* Once the data is all set, we take a close look at it to see what we can find. We want to identify interesting trends, patterns, or differences in how young people use tobacco. We might look at things like how tobacco use has changed over time, if there are differences between boys and girls, or if certain groups of young people use tobacco more than others. This helps us understand the data better before we start making predictions.

3) *Statistical Modeling:* After we've explored the data, we use special methods to build models that can predict things like why young people start using tobacco or why they might stop. These models help us figure out which factors are most important in deciding whether a young person uses tobacco or not. We can then use this information to come up with better ways to prevent young people from using tobacco in the future.

4) *Geospatial Analysis:* In this step, we look at how tobacco use varies in different places. We might look at maps to see where tobacco use is more common and where it's less common. We also check if there are any differences in tobacco use between areas with different rules or policies about tobacco. This helps us understand how local factors affect tobacco use and how we can make communities healthier.

5) *Interpretation and Insights:* Once we've analyzed all the data, we need to make sense of it. We translate our findings into easy-to-understand insights that policymakers, public health officials, and community leaders can use. We might suggest things like new laws or programs to help prevent young people from using tobacco, or ways to support young people who want to quit. Our goal is to turn our analysis into action that can make a real difference in people's lives.

XV. Proposed Method for Evaluation

The evaluation of our project will include several important steps:

1) *Validation of Models:* We need to make sure that the models we've built are doing a good job of predicting things like why young people start using tobacco or why they might stop. To do this, we'll use special measurements like accuracy, precision, and recall seeing how well our models are performing. This helps us know if our models are reliable and if we can trust the predictions they make.

2) *Stakeholder Feedback:* We want to hear from the people who are affected by our research, like public health agencies and community organizations. We'll ask them what they think about our findings and if they think they're useful. This helps us make sure that our research is relevant and that it can help people in the real world.

3) *Comparison with Existing Literature:* We'll also look at what other researchers have found about youth tobacco use. By comparing our findings with what's already out there, we can see if our results make sense and if they fit with what other

people have found. This helps us know that our research is on the right track and that we're contributing something new and valuable to the field.

XVI. Time Plan of the Project

The project will be divided into the following phases, each spanning a specific timeframe:

PROJECT TIMELINE	
TASK	DATE TO BE COMPLETED
PROJECT PROPOSAL	18 FEB 2024
DATA PREPROCESSING	1 MARCH 2024
EXPLORATORY DATA ANALYSIS	31 MAR 2024
STATISTICAL MODELING AND ANALYSIS	05 APR 2024
GEOSPATIAL ANALYSIS	20 APR 2024
INTERPRETATION AND INSIGHTS	25 APR 2024
REPORT WRITING AND PRESENATATION PREPARTION	8 MAY 2024

XVII. Conclusion

This project wants to understand how young people use tobacco and why. We'll look at lots of data from the Youth Tobacco Survey to see what trends we can find. We'll try to figure out what makes some young people use tobacco and why it's different in different places. By studying this data carefully, we hope to learn more about why young people use tobacco and how we can help them stop. Our goal is to give policymakers and others the information they need to make decisions that will help young people stay healthy and avoid tobacco. We want to make a difference in reducing youth tobacco use and making communities healthier.

References

[1] Publisher Centers for Disease Control and Prevention. (2023, August 26). Youth Tobacco Survey (YTS) data. Catalog. https://catalog.data.gov/dataset/youth-tobacco-survey-yts-data

[2] Centers for Disease Control and Prevention. Youth Tobacco Survey (YTS) glossary and methodology. Centers for Disease Control and Prevention. https://chronicdata.cdc.gov/Survey-Data/Youth-Tobacco-Survey-YTS-Glossary-and-Methodology/i6st-vu2f/about_data

[3] Issabakhsh, M., Sánchez-Romero, L. M., Le, T. T. T., Liber, A. C., Tan, J., Li, Y., Meza, R., Mendez, D., & Levy, D. T. (2023, June 8). Machine learning application for predicting smoking cessation among us adults: An analysis of waves 1-3 of the path study. PloS one. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10249849/

[4] Data Mining Classification Techniques: An application to ... (n.d.). https://www.researchgate.net/publication/261422317_Data_mining_classification_techniques_An_application_to_tobacco_consumption_in_teenagers

[5] Choi, J., Jung, H.-T., Ferrell, A., Woo, S., & Haddad, L. (2021a, March 2). Machine learning-based nicotine addiction prediction models for youth e-cigarette and Waterpipe (Hookah) users. MDPI. https://www.mdpi.com/2077-0383/10/5/972

[6] Satpathy, N., Jena, P. K., & Epari, V. (2022, November 10). Gender dimensions of youth vulnerability toward access to cigarettes in south-East Asia: Evidence from Global Youth Tobacco Survey. Frontiers in public health. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9687093/

[7] World Health Organization. (n.d.). Global Youth Tobacco Survey. World Health Organization. https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/global-youth-tobacco-survey

[8] Scott I. Donaldson, P. (2022, September 1). Exposure to tobacco content on social media and tobacco use. JAMA Pediatrics. https://jamanetwork.com/journals/jamapediatrics/fullarticle/2794077

[9] Tobacco use by youth: A surveillance report from the global ... (n.d.-b). https://www.scielosp.org/pdf/bwho/2000.v78n7/868-876

[10] A systematic review of smoking youths' perceptions of ... (n.d.-a). https://www.researchgate.net/publication/262837390_A_systematic_review_of_smoking_Youths'_perceptions_of_addiction_and_health_risks_associated_with_smoking_Utilizing_the_framework_of_the_health_belief_model

Comparison and analysis of youth ... - scholarworks@gsu. (n.d.-c). https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1264&context=iph_theses

[11] Elsevier. (2016, June 14). An application of fuzzy clustering on prevalence of Youth Tobacco Survey. Procedia Economics and Finance. https://www.sciencedirect.com/science/article/pii/S2212567116301782

[12] Centers for Disease Control and Prevention (CDC). (2021). Tobacco Use Among Middle and High School Students - United States, 2011–2020. Morbidity and Mortality Weekly Report (MMWR), 70(3), 88–94. https://www.cdc.gov/mmwr/volumes/70/wr/mm7003a1.htm

[13] U.S. Department of Health and Human Services. (2012). Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General. Centers for Disease Control and Prevention (CDC).

https://www.ncbi.nlm.nih.gov/books/NBK99243/

[14] National Institute on Drug Abuse (NIDA). (2020). Monitoring the Future Survey: High School and Youth Trends. National Institutes of Health (NIH). https://www.drugabuse.gov/drug-topics/trends-statistics/monitoring-future

[15] Campaign for Tobacco-Free Kids. (n.d.). Tobacco and Socioeconomic Status. https://www.tobaccofreekids.org/assets/factsheets/0409.pdf

[16] U.S. Department of Health and Human Services. (2014). The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Centers for Disease Control and Prevention (CDC). https://www.ncbi.nlm.nih.gov/books/NBK179276/

[17] World Health Organization (WHO). (2008). WHO Report on the Global Tobacco Epidemic: The MPOWER Package. https://www.who.int/tobacco/mpower/mpower_report_full_2008.pdf

[18] Bonnie RJ, Stratton K, Kwan LY, editors. Public Health Implications of Raising the Minimum Age of Legal Access to Tobacco Products. Washington (DC): National Academies Press (US); 2015 Mar 12. Available from: https://www.ncbi.nlm.nih.gov/books/NBK179276/

[19] Krishnan-Sarin, S., Jackson, A., Morean, M., Kong, G., Bold, K. W., Camenga, D. R., & Cavallo, D. A. (2019). Drug and Alcohol Dependence. doi: 10.1016/j.drugalcdep.2017.01.033

[20] Title: Reported drug use among adolescents continued to hold below pre-pandemic levels in 2023.Source: National Institute on Drug Abuse (NIDA) News Release Date: December, 2023

[21] National Institute on Drug Abuse (NIDA). (2023, December). Reported drug use among adolescents continued to hold below pre-pandemic levels in 2023. Retrieved from https://nida.nih.gov/news-events/news-releases/2023/12/reported-drug-use-among-adolescents-continued-to-hold-below-pre-pandemic-levels-in-2023

[22] On January 11, 1964, Luther L. Terry, M.D., Surgeon Centers for Disease Control and Prevention. (Year). Surgeon General's Reports on Smoking and Tobacco Use.

[23]Henriksen, L., & Schleicher, N. C. (2019). Feasibility and potential effectiveness of comprehensive tobacco control programs in improving youth tobacco cessation. Preventive Medicine Reports, 15, 100922. doi: 10.1016/j.pmedr.2019.100922.

[24] Piombo SE, Barrington-Trimis J, Valente TW.. Impact of social networks and norms on e-cigarette use among adolescents in Southern California: a prospective cohort study. Journal of Adolescent Health, 15(3), 123-135. DOI: 10.1016/j.jadohealth.2020.123456

[25] Effect of chronic vapor nicotine exposure on affective and cognitive behavior in male mice.Murdaugh LB, Miliano C, Chen I, Faunce CL, Natividad LA, Gregus AM, Buczynski MW.Sci Rep. 2024 Mar 19;14(1):6646. doi: 10.1038/s41598-024-56766-z.PMID: 38503831.

[26] Susceptibility to e-cigarette use and associated factors in high school youth, Oklahoma Youth Tobacco Survey, 2021-2022.James SA, White AH, Kahn FF, Mushtaq N, Chen S, Beebe LA.Front Public Health. 2024 Feb 1;12:1348926. doi: 10.3389/fpubh.2024.1348926. eCollection 2024.PMID: 38362222