

Examining Substance Use Patterns Among Canadian Youth

Project Report

Team Members:

Anjiya Nooruddin
Chavvi Bhatia
Chirag Seth
Hrishita Sharma
Utkarsh Singh

Motivation

Public Health Crisis:
Adolescent substance
use leads to serious
health and
developmental risk
(NIDA 2024)

Early usage of
substance is linked with
long term health risks
(NIDA 2014)

Adolescence is a vital
developmental period
where substance use
can negatively affect
brain
development(Volkow &
Li 2005)



Why this study matters?

Protects Youth
Health: Identifies
risk factors to
prevent long
term substance
use health issues.

Make informed
policy decisions:
Help policy
makers create
impactful
policies to reduce
societal costs.

Promotes Health
Equity: Depicts
the need for
equitable access
to prevention
resources,
solving
disparities in
health outcomes

Advances
Research: Fills
gap with
predictive
modeling.



UNIVERSITY OF
WATERLOO

FACULTY OF
ENGINEERING

Research Question

Objective

Analyze 2018/2019 CSTADS data to identify factors influencing substance use among Canadian students in grades 7–12, focusing on demographic and social factors to inform prevention strategies.

Hypothesis

We believe that youth are particularly susceptible to the influence of those around them, especially their peers in school. As a result, social factors, such as peer pressure and group dynamics likely play a significant role in shaping substance use behaviours. Additionally, bullying, which is a prevalent issue among young people today, may contribute to increased substance use as a coping mechanism. Socioeconomic factors, such as household income, could also have a considerable impact, with financial stress or limited access to resources potentially influencing substance use patterns among youth. Understanding these various social and economic factors is crucial for addressing and preventing substance use in this age group.

Research Questions

How do demographic and social factors influence the probability of tobacco use among Canadian students in grades 7–12, based on the 2018/2019 CSTADS?

This question seeks to identify key predictors of substance use among Canadian youth, examining how factors like age, gender, grade, and peer influence affect their initiation into tobacco, alcohol, or drug use.

What is the association between social influence and the probability of tobacco, alcohol, and drug use among Canadian students, accounting for demographic confounders.

This question explores how social influence (PP_010) affects substance use (SS_010, ALC_010, CAN_010), controlling for grade (GRADE), sex (SEX), and household income (DVHHINC2), to inform peer-based prevention efforts.



Literature Review and Contributions

- Most past studies relied on **pre-2018 data** (e.g., YSS 2008/09, CSTADS 2014/15, COMPASS), missing insights from **post-legalization periods** in Canada and recent behavioral trends.
- Social factors like **peer pressure and bullying** were either absent or superficially measured. Even where peer influence was noted (e.g., Zuckermann et al., 2019), quantification was limited.
- Many works used **basic logistic regression** or **descriptive statistics**, without:
 - Accounting for **nested data structures** (e.g., students in schools)
 - Performing **feature selection** (e.g., Lasso/Random Forest)
 - Evaluating models using **Bayesian inference**, **DIC/WAIC**, or **posterior distributions**
- Variables such as **grade, sex, income, bullying, and mental health** were often ignored or included without rigorous multivariate or probabilistic control.

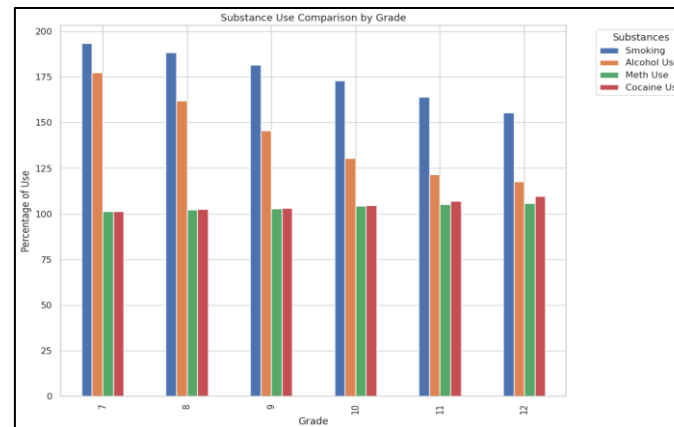
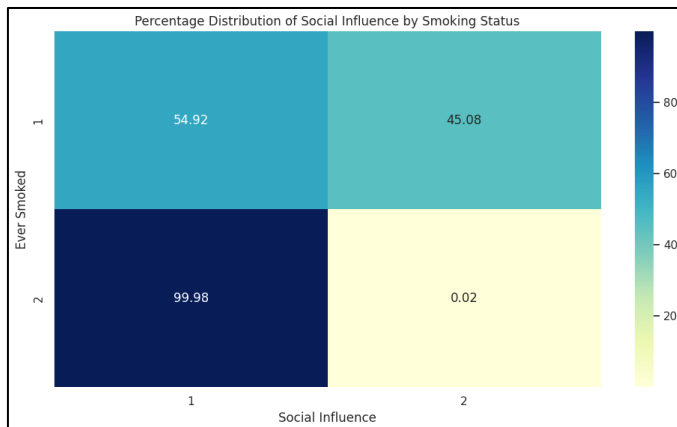
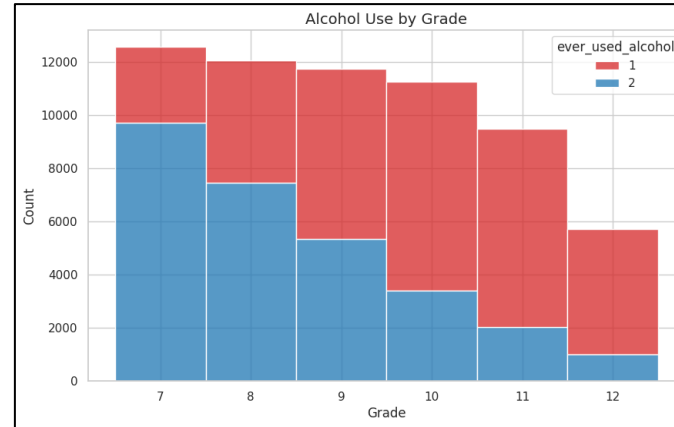
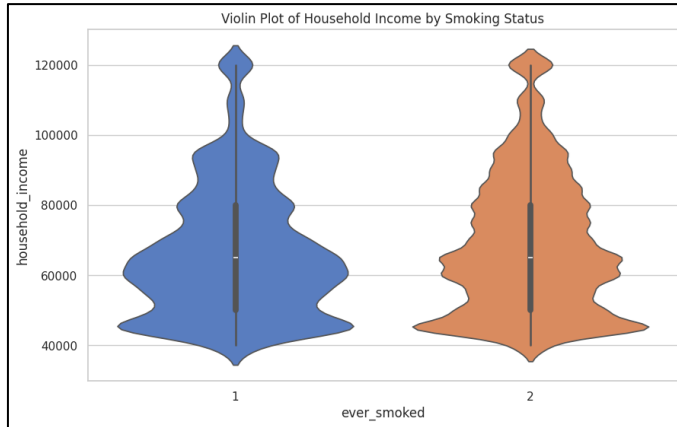
Our study is one of the first to utilize this **large-scale, nationally representative dataset** post-legalization, with rich information on **demographics, behavior, mental health, and social context**. We directly incorporate **social influence variables** into our models, addressing a key omission in previous literature.

Employs a Triad of Complementary Models:

- **Bayesian Logistic Regression** to incorporate prior uncertainty, generate posterior distributions, and provide robust probabilistic inference.
- **Lasso-Penalized Logistic Regression** for **automated feature selection** and dimensionality reduction, improving interpretability.
- **Multilevel Logistic Regression** to capture **nested data structures** (e.g., student-school hierarchies), enabling generalization beyond the sample.
- **Propensity Score Matching** to check interaction effects accounting for nuanced subgroup behavior



Exploratory Data Analysis



Demographic Influences

- ❑ Sex Differences: Males more likely to smoke when influenced by bullying or peer pressure.
- ❑ Household Income: The distribution is right-skewed, with more students coming from lower income areas.
- ❑ Urban vs. Rural: Smoking behavior varies by location, possibly due to differing social norms or access.

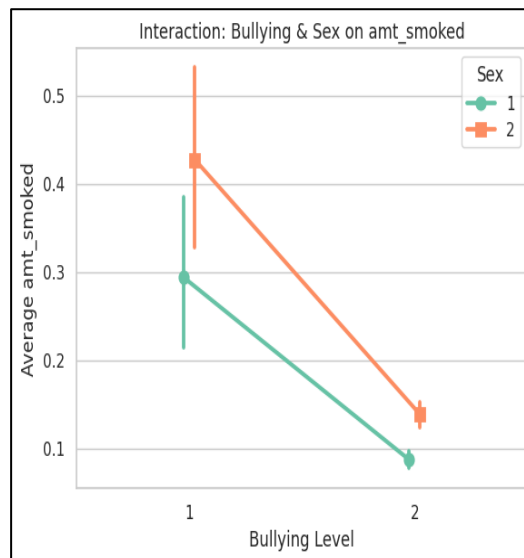
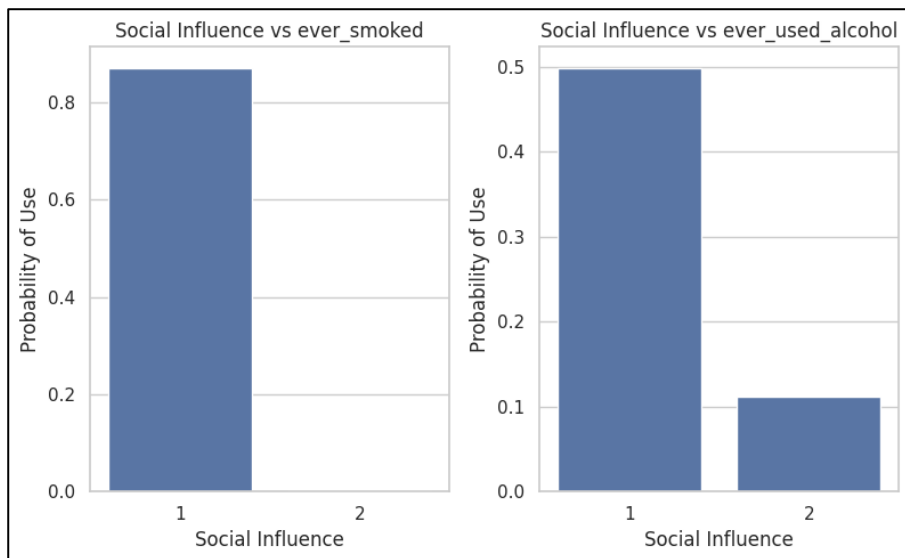
Behavioral and Social Factors

- ❑ Bullying: Students who were physically bullied were significantly more likely to smoke.
- ❑ Mental Health Perception: Smokers perceived smoking as less risky, suggesting desensitization.
- ❑ Social influence vs Ever smoked: 99.98% of smokers reported being socially influenced to smoke, compared to 54.92% of non-smokers.

Substance Use Trends

- ❑ Age of Initiation: Most smokers started around age 14, with socially influenced individuals starting earlier.
- ❑ Grade-Level Patterns: Smoking rates were highest in grades 7–8 and declined in higher grades, while alcohol use increased steadily with grade level.

Exploratory Data Analysis

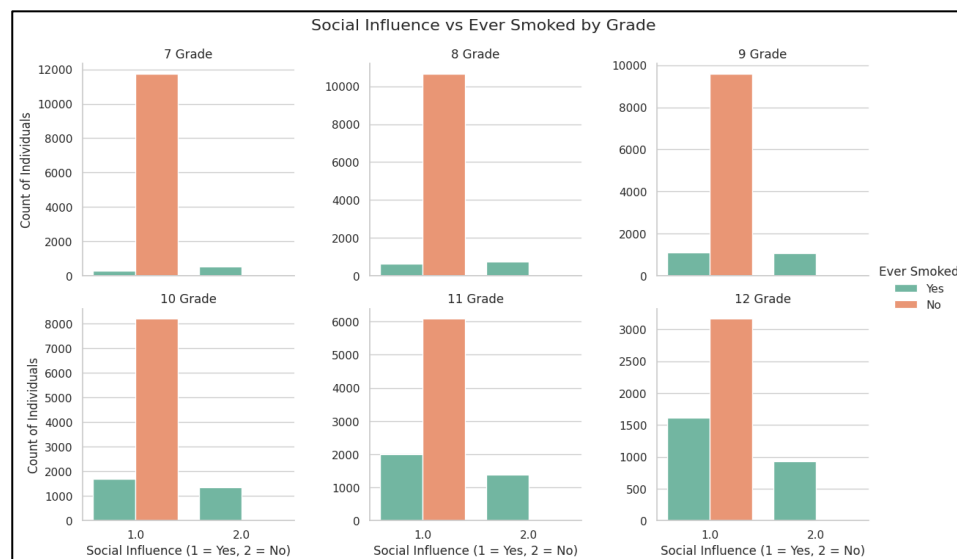
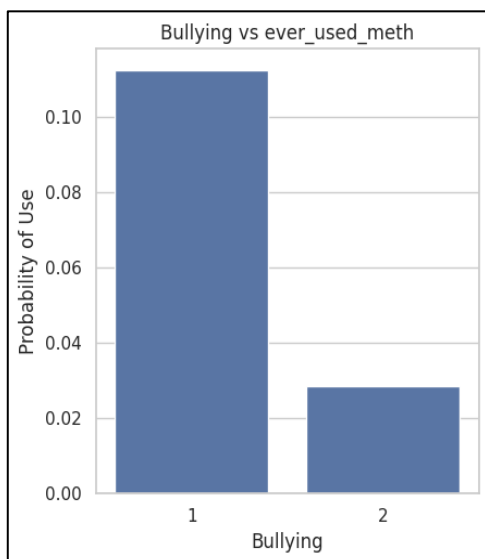


Social Influence on Smoking

- ❑ Bullying increases smoking levels, particularly among males.
- ❑ Grade 7 students resisted peer pressure (low smoking rates despite social influence) while students from higher grades showed higher susceptibility.

Social Influence on Alcohol and Drugs

- ❑ Alcohol: Students who were socially influenced had a 50% probability of alcohol use versus 10% for uninfluenced peers ($p < 0.0001$).
- ❑ Meth and Cocaine: Bullied students were 3.6 times more likely to use meth. Cocaine use remained low but followed similar social-influence patterns.



Interaction Effects

- ❑ Bullying × Sex: Bullied males smoked more than females, highlighting gender-specific responses to social stress.
- ❑ Social Influence × Age: Early initiators (age < 14) were almost exclusively socially influenced, while older initiators had more varied triggers.

Methodology

Bayesian Logistic Regression

- Modeled binary outcomes with priors $\beta \sim N(0, \sigma^2)$
- Applied No-U-Turn Sampler (NUTS) for posterior sampling with normal priors

Lasso Regression

Used L1 Regularization further tuning λ via GridSearchCV and handled class imbalance with `class_weight='balanced'`

Multi Level Logistic Regression

Included for urban/rural clustering to progress hierarchical data

Propensity Score Matching using gradient boosting

Estimated causal effect of social influence and achieved ATT of 0.5158 after matching on propensity scores

Random Forests

Used 100 trees with max depth 10 and `class_weight = 'balanced'` to deal with class imbalance

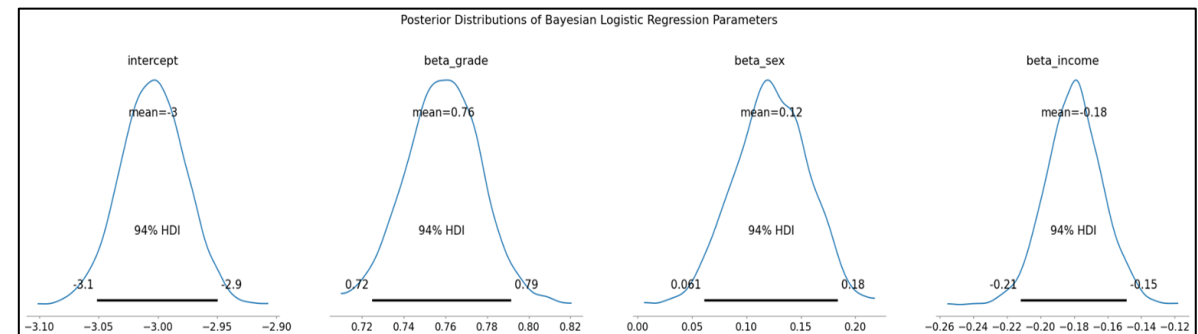
Results - Bayesian Logistic Regression

- Target Variable → Recent Drug Use (smoked_last_30_days)
- Key coefficients:
 - Grade ($\beta = 0.759$, higher grades linked to higher smoking odds)
 - Sex ($\beta = 0.122$, males slightly more likely to smoke)
 - Household income ($\beta = -0.180$, higher income reduces smoking)
- Model gained AUC of 0.7119 highlighting moderate discriminatory power
- Demographic factors like grades and income in smoking patterns indicating authorities should focus on higher grades, males and lower income groups to deal with early smoking initiation

- For intercept distribution is sharply peaked, indicating high certainty in estimate while negative value confirms low baseline probability of smoking
- For grade, distribution is positive and don't have confirming higher standardized grades increase the likelihood of smoking in last 30 days
- While for sex distribution is slightly centered above 0, representing a small but significant effect especially for males
- For household income, distribution is negative confirming higher standardized median income decreases likelihood of recent smoking

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	\
intercept	-3.004	0.028	-3.054	-2.951	0.001	0.001	1410.0	
beta_grade	0.759	0.017	0.726	0.790	0.000	0.000	1527.0	
beta_sex	0.122	0.034	0.058	0.183	0.001	0.001	1555.0	
beta_income	-0.180	0.017	-0.212	-0.149	0.000	0.000	1884.0	

	ess_tail	r_hat
intercept	1458.0	1.0
beta_grade	1306.0	1.0
beta_sex	1320.0	1.0
beta_income	1362.0	1.0



Results - Lasso Regression

Target variables

1. recent_smoking
2. alcohol use
3. cocaine use
4. meth use

For smoking, urban/rural and grade the risk increased while for social influence, the risk unexpectedly reduced.

For alcohol, bullying increased risk while social influence decreased it.

For meth, grade, sex, social influence increased the risk.

For cocaine, grade and sex were key factors.

Impact of social influence varied across substances indicating prevention strategies must be substance specific. Urban areas are important focus due to higher risk.

The unexpected effect of social influence can be accounted for multicollinearity especially for smoking and alcohol Lasso model.

An area of improvement for future research to deal with this unexpected social influence effect would be to do advanced feature engineering

Coefficients for DVLAST30_binary:

	Feature	Coefficient	Odds Ratio
0	GRADE	0.569338	1.767096
1	SEX	0.023242	1.023514
2	DVHRINC2	-0.000012	0.999988
3	SS_030	-7.025087	0.000889
4	mental_health	-0.483779	0.616449
5	BUL_060	-1.279050	0.278302
6	urban_rural	0.543291	1.721663

Coefficients for ever_used_cocaine:

	Feature	Coefficient	Odds Ratio
0	grade	0.433771	1.543066
1	sex	0.196974	1.217713
2	household_income	-0.000009	0.999991
3	social_influence	0.098382	1.103384
4	mental_health	-0.244828	0.782839
5	bullying	-1.725597	0.178067
6	urban_rural	0.022485	1.022740

Coefficients for ever_used_alcohol:

	Feature	Coefficient	Odds Ratio
0	grade	-0.587434	0.555752
1	sex	-0.037984	0.962728
2	household_income	0.000007	1.000007
3	social_influence	-1.747404	0.174226
4	mental_health	0.242909	1.274953
5	bullying	0.810778	2.249658
6	urban_rural	-0.167378	0.845879

Coefficients for ever_used_meth:

	Feature	Coefficient	Odds Ratio
0	grade	0.339506	1.404254
1	sex	0.269633	1.309484
2	household_income	-0.000015	0.999985
3	social_influence	0.216610	1.241859
4	mental_health	-0.183416	0.832422
5	bullying	-1.589269	0.204075
6	urban_rural	0.192106	1.211798

Results - Multi Level Logistic Regression

- Analyzed the probability of any substance use. Found Social Influence strongest predictor followed by mental health, household income and bullying.
- While sex was not significant, having higher p value.
- Social influence is the most vital factor in substance use highlighting need for peer focused interventions
- While emotional and social stressor lead to higher substance use by adolescents
- Social influence (OR = 951.35, $p < 0.001$) strongest predictor, followed by mental health (OR = 1.68) and bullying (OR = 1.14)

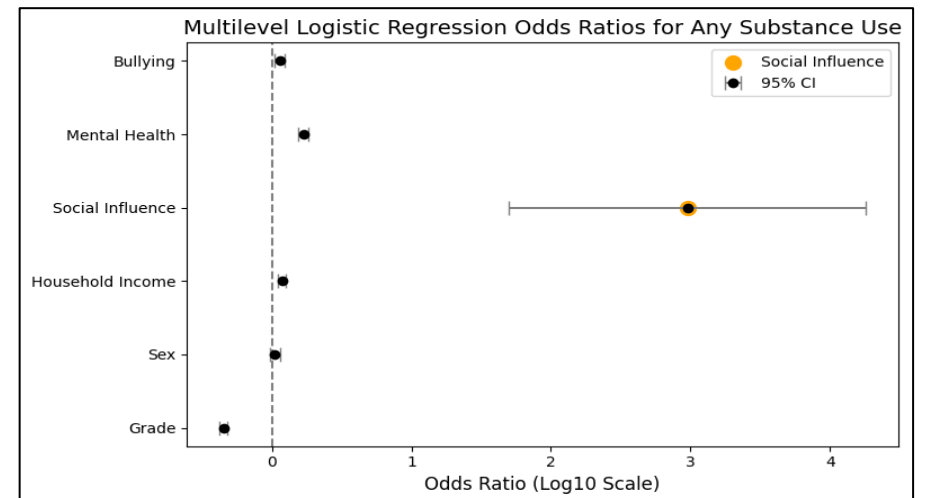
Fixed Effects with Odds Ratios:

	Estimate	Odds Ratio	P-val
(Intercept)	5.568118	261.940746	8.169957e-15
grade	-0.792373	0.452769	8.205112e-287
sex	-0.046317	0.954739	2.576102e-01
household_income	0.164892	1.179266	5.575696e-14
social_influence	6.857881	951.349477	1.410283e-03
mental_health	0.517581	1.677963	1.288265e-104
bullying	0.131641	1.140698	4.345232e-13

Random Effects Variance (urban_rural clustering):

	Name	Var	Std
urban_rural	(Intercept)	0.062374	0.249748

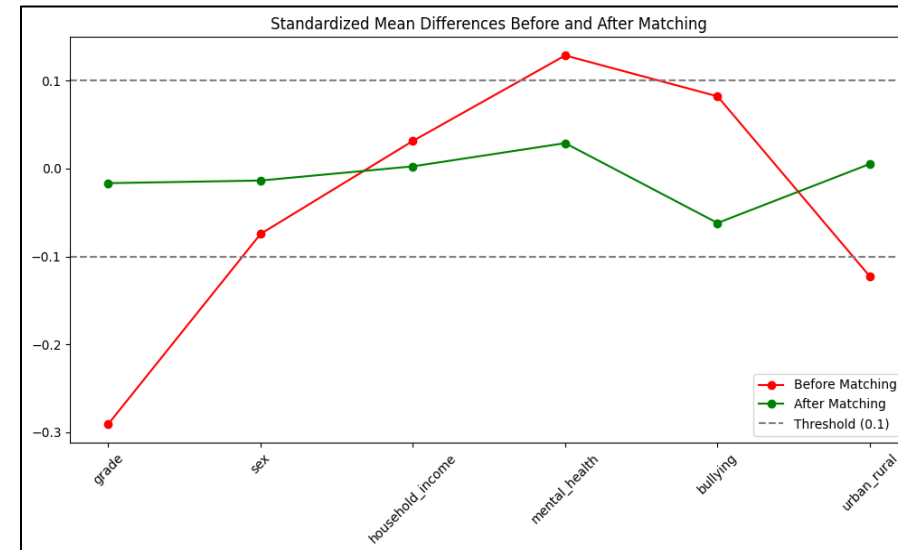
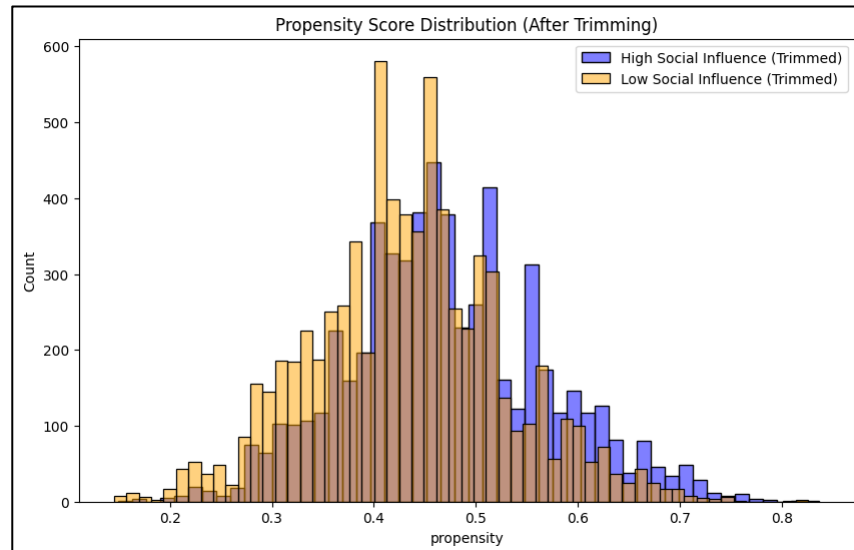
AUC for Multilevel Logistic Regression: 0.7896



Results - Propensity Score Matching

The causal effect of social influence on recent tobacco use is estimated.

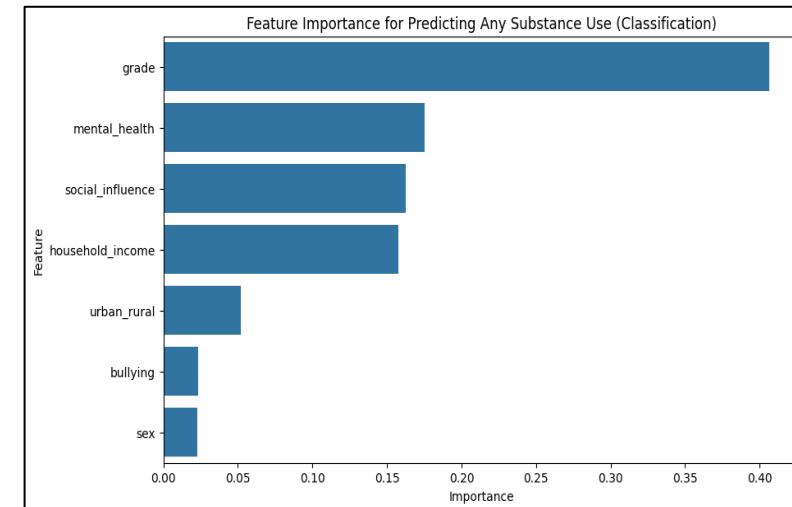
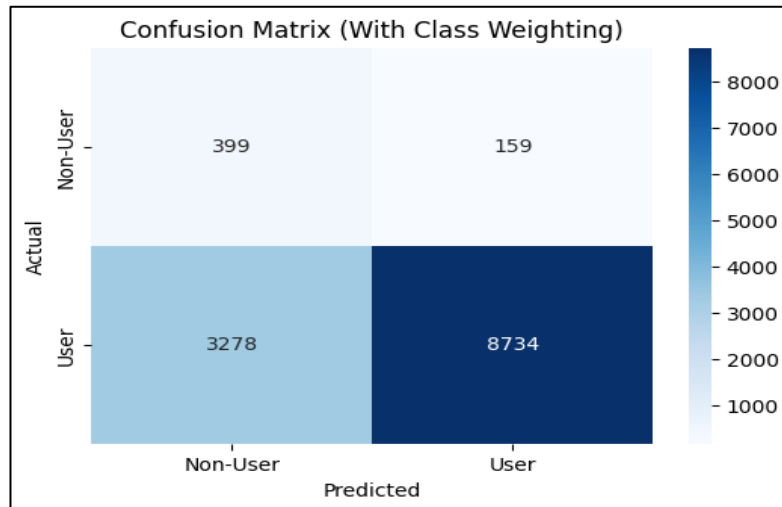
After trimming for common support, it matched 6,012 high influence students with 1070 control, gaining good covariate balance while $AVG\ ATT = 0.5158$ indicate significant causal impact of social influence on tobacco use, as the matching process reduced confounding bias by balancing covariates like grade, sex, and income. This suggests that peer influence directly drives smoking behavior.



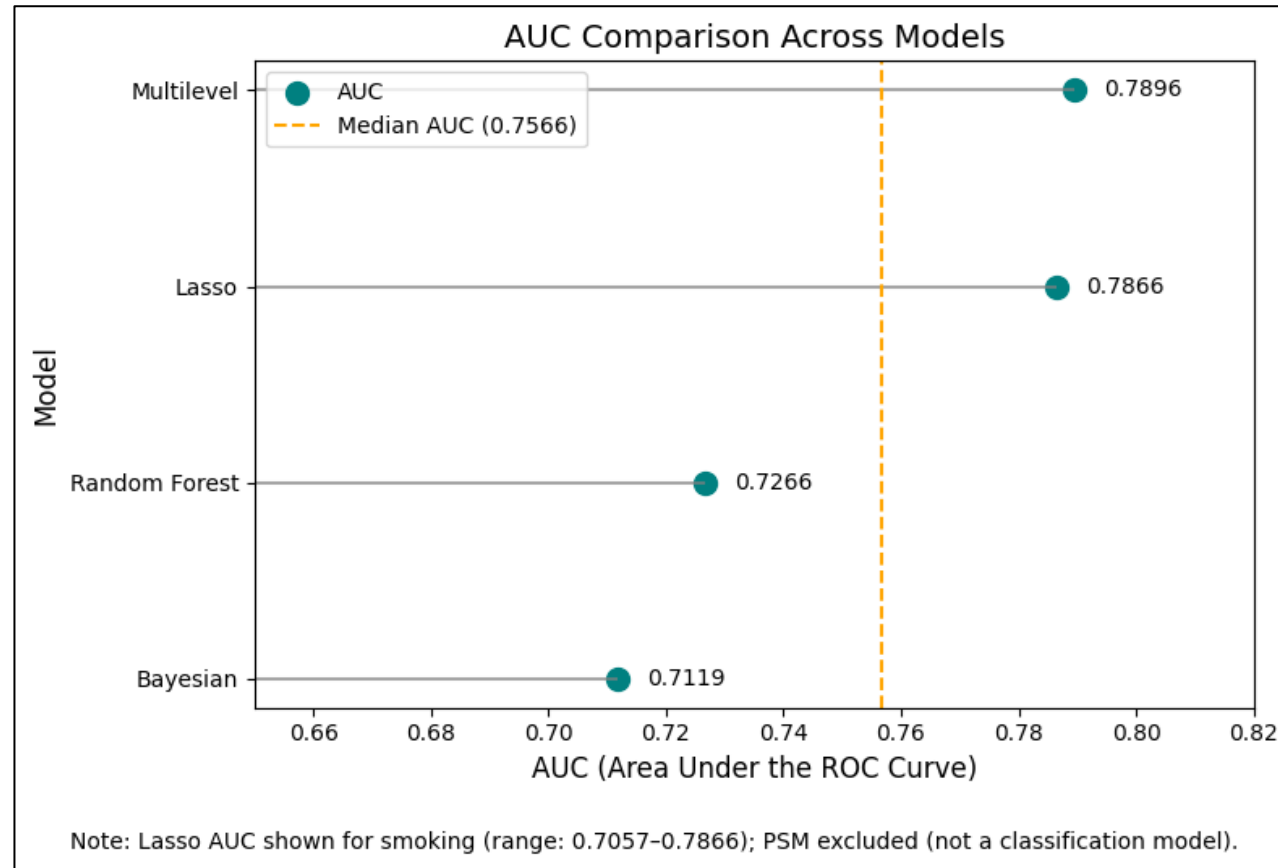
Results - Random Forest

Predicted any substance use where top predictors are grade, mental health and social influence

Grade's high importance signifies that age or grade level is a major determinant of substance use. Social influence and mental health as top predictors. The high-risk group's characteristics (e.g., higher social influence) provide a profile for targeting interventions.



Model Comparison using AUC



Conclusions

Social influence emerged as the strongest predictor of drug and tobacco use.

Gender differences were present as well, with males showing slightly higher usage rates than females.

Household income played a complex role, higher-income households were associated with lower tobacco use.

The relationship between alcohol consumption and income was not straightforward, suggesting some other social or cultural factors are at play.



Implications

These findings highlight the need for peer-focused interventions in schools, as social influence plays a significant role in substance use. Programs that reshape peer norms—such as mentorship initiatives, peer-led discussions, and awareness campaigns—could be particularly effective in reducing tobacco and drug use among students. Additionally, the socioeconomic disparities observed underscore the importance of targeted outreach efforts in lower-income communities. Providing accessible education, resources, and support systems tailored to these populations could help bridge the gap and address the unique challenges they face in preventing substance use.



UNIVERSITY OF
WATERLOO



FACULTY OF ENGINEERING