

# Project Overview



**Problem Statement:** Predict heroin usage / addiction.

**Proposed Solution:** Apply machine learning models to analyze substance use and demographic data.

**Potential Impact:** Inform targeted interventions and public health strategies.

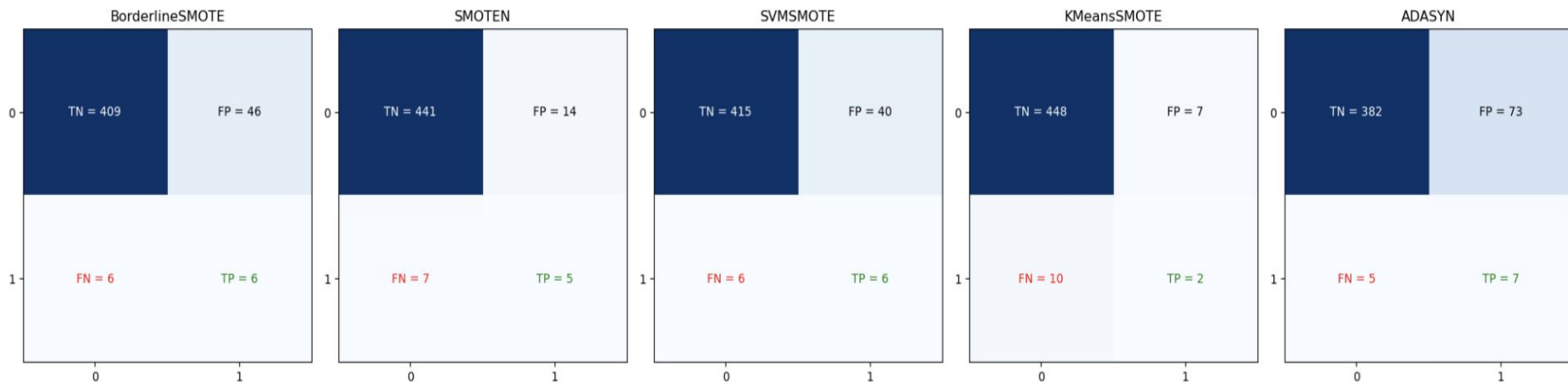
# Models Comparison

Target : Heroin usage

	Hyperparameter	Training Accuracy	Test Accuracy	Notes
Logistic regression	none	97.85%	97.85%	good performance but identifying non-users cor...
BorderlineSMOTE	C=1	93%	89%	balanced performance with good test accuracy
SMOTEN	C=1	96%	96%	excellent performance, best among the variations
SVM SMOTE	C=10	94%	90%	good balance of precision and recall
KMeans SMOTE	C=0.1	nan	96%	handled imbalanced data well, though cross-val...
ADASYN	C=10	88%	83%	focused on difficult cases, moderate test accu...
Decision Tree (Base)	None	100%	95.93%	performance is excellent as expected
Decision Tree	max_depth=5, min_impurity_decrease=0.0, min_sa...	97%	97%	high cross-validation and test accuracy
Random Forest	max_depth=20, min_samples_leaf=2, min_samples_...	97%	97%	high cross-validation and test accuracy

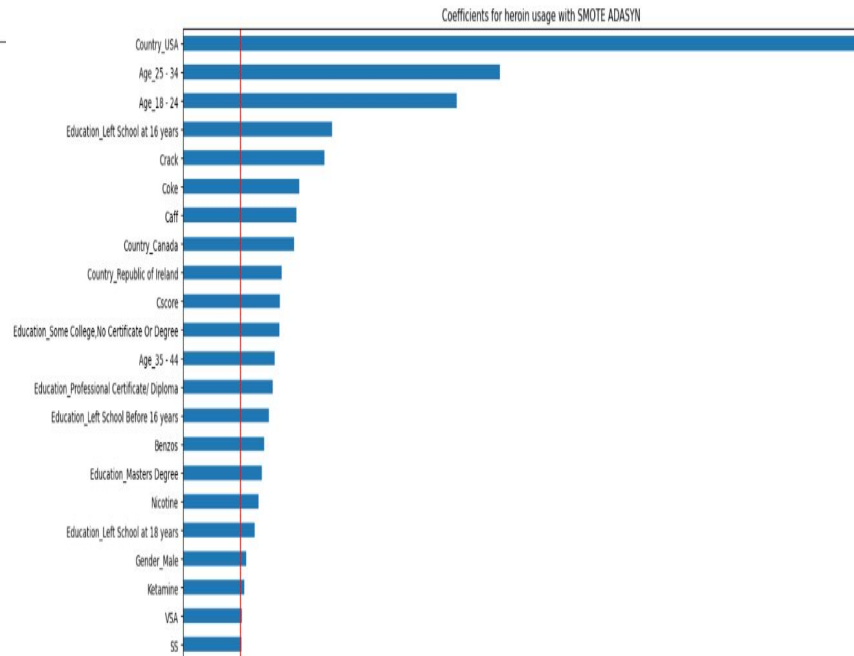
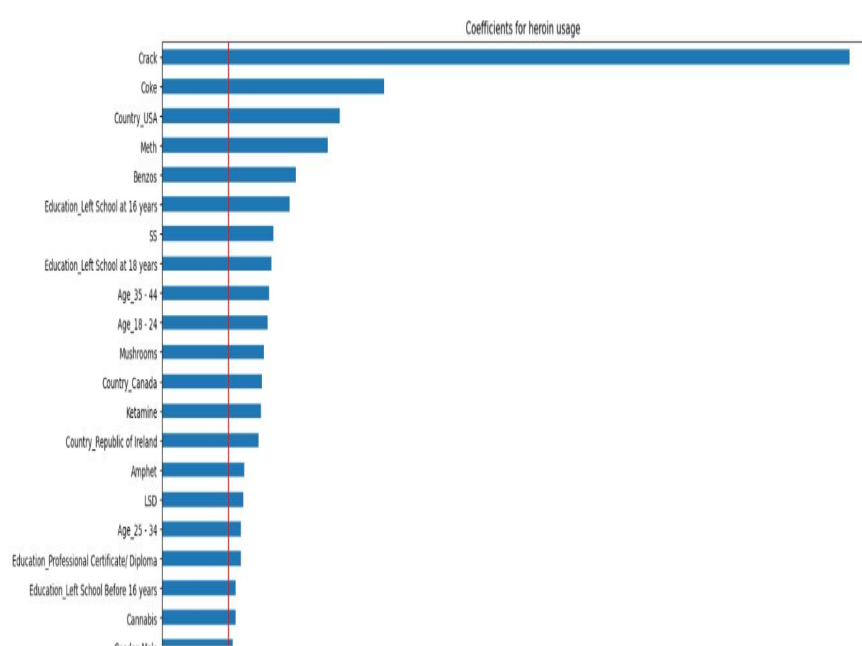
# Evaluation Metrics comparison

- ❖ **ADASYN** (Adaptive Synthetic Sampling) and **SVM SMOTE** (Support Vector Machine and Synthetic Minority Over-sampling Technique) are the most effective methods for identifying individuals likely to get addicted to heroin due to their higher true positive rates.



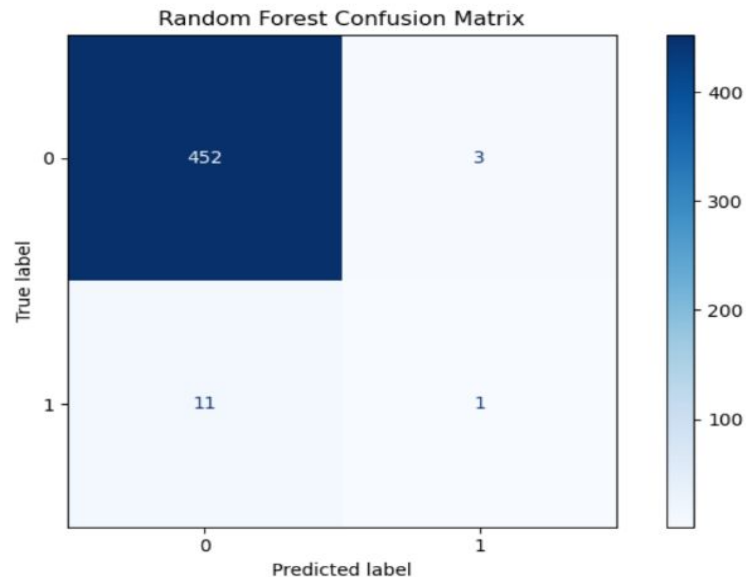
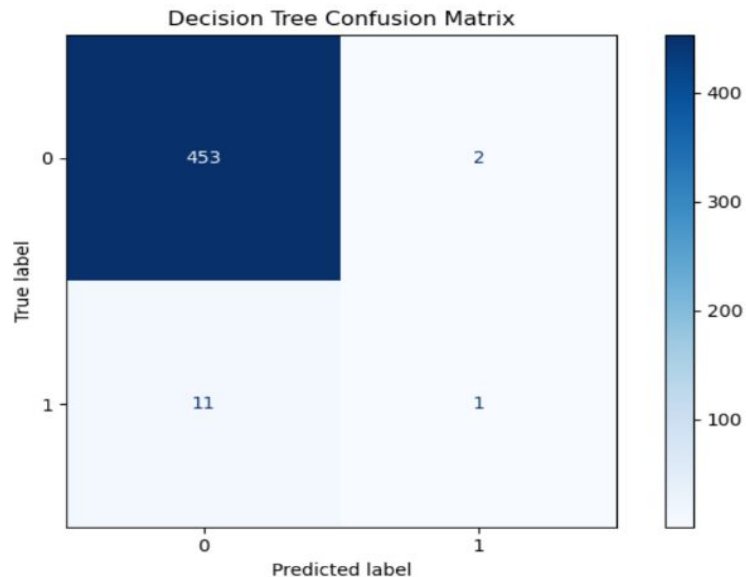
# Coefficient Analysis

- ❖ **Without Hyperparameter Tuning:** High coefficients like **Crack (10.37)**, **Coke (3.49)** and **USA (2.68)** indicate strong associations.
- ❖ **With ADASYN (Adaptive Synthetic Sampling) SMOTE:** Coefficients are balanced; **USA (11.90)** , **Age 25-34 (5.52)** and **Age 18-24 (4.77)** stand out.



# Decision Tree vs Random Forest Evaluation

- ❖ Decision Tree and Random Forest demonstrate consistent performance, each correctly identifying 1 true positive, alongside achieving 97% test accuracy for negative cases.



## Next Steps

- ❖ Fit different drugs as a target variable using the same pipeline
- ❖ Compare the models and results
- ❖ Coefficient analysis
- ❖ Feature Selection analysis based on the results
- ❖ Focus on Personality type analysis

**Thank you!!!**