

# **Substance Classification**

**Jacob Levine**

# Problem

- Given information regarding an individual's personality traits, can we predict whether they are at risk for using marijuana?
- Other substances?

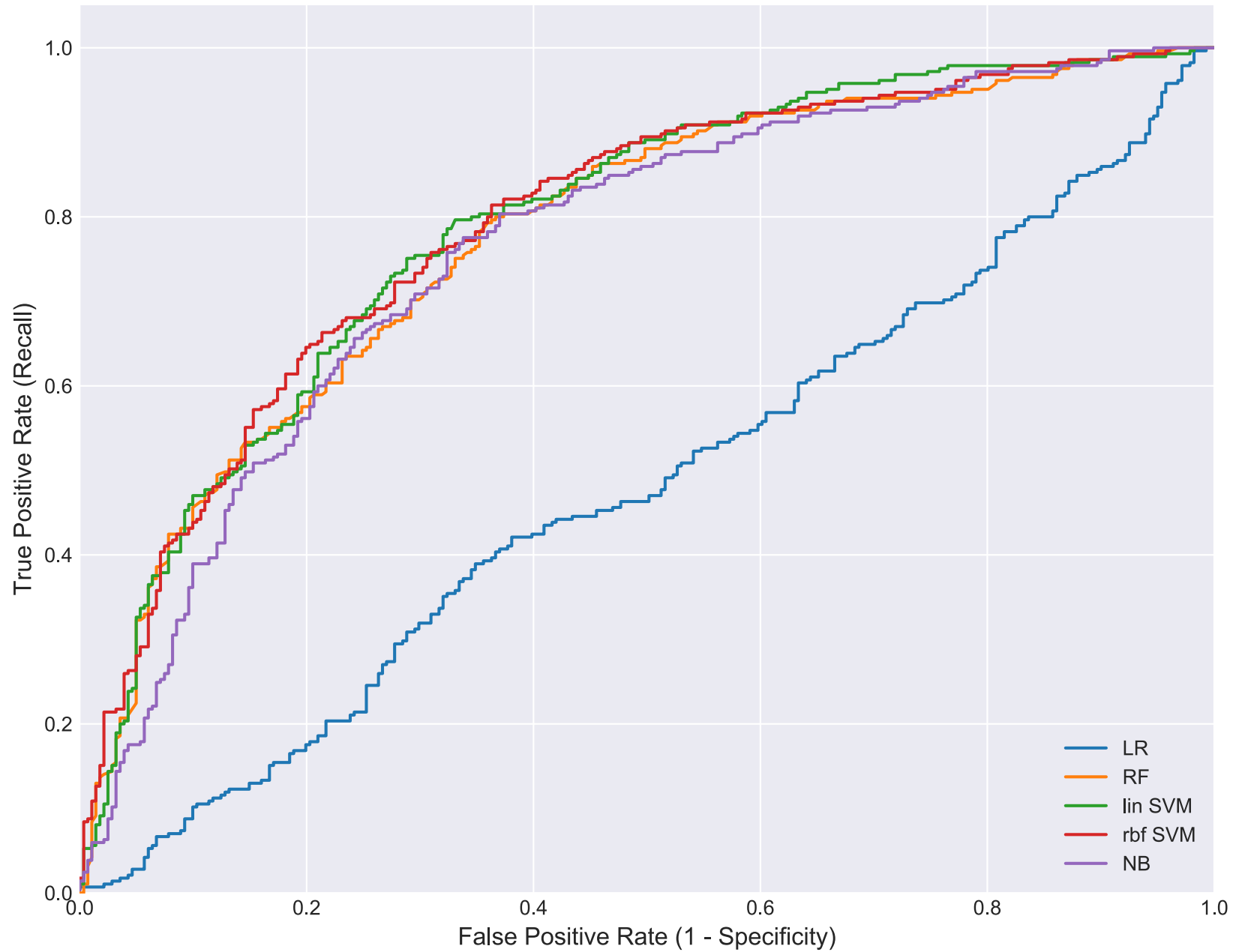
# Data

- **Source**
  - UCI Machine Learning Repository
- **Training Set:**
  - 1,319 Observations
  - 0 (Non-User) – 608 Observations
  - 1 (User) – 711 Observations

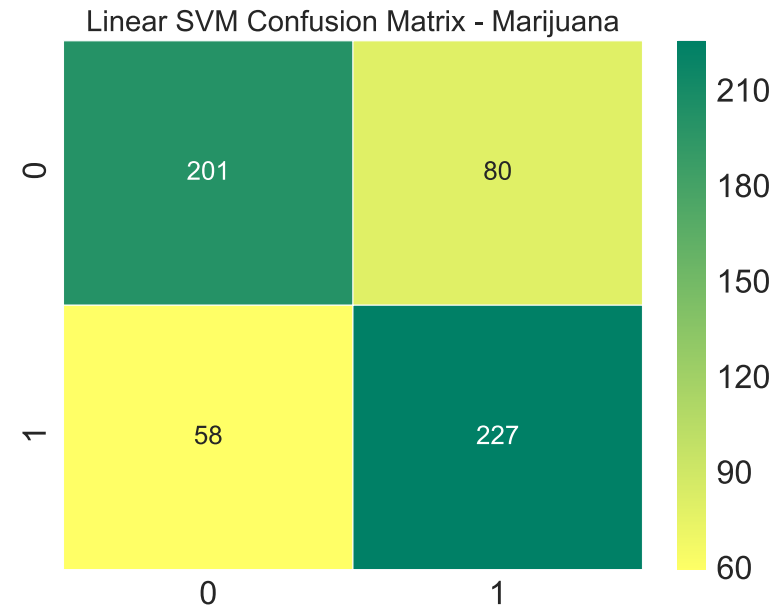
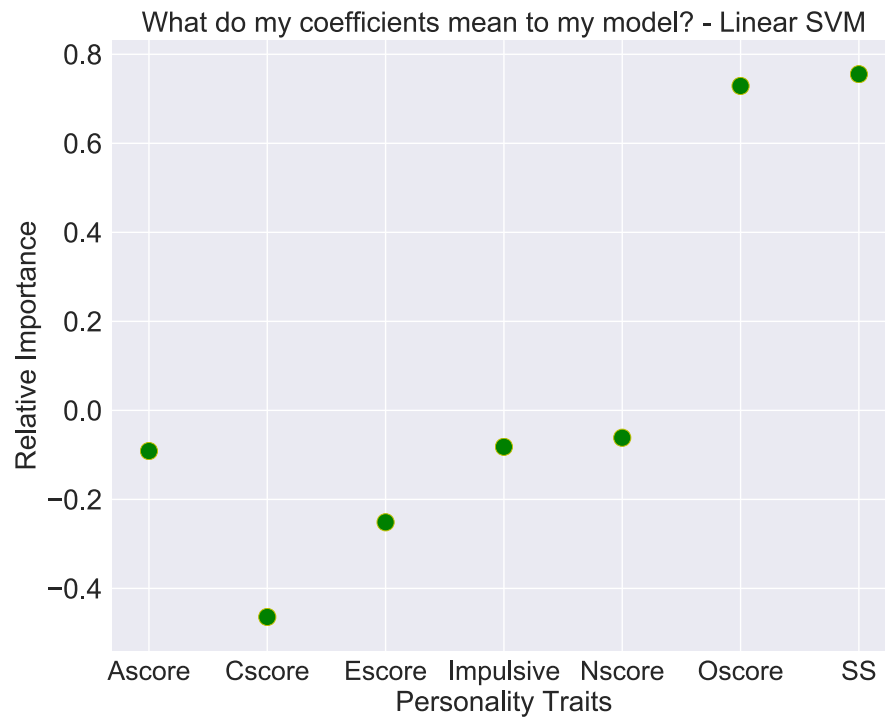
# NEO-FFI-R Framework

- Nscore – Neuroticism
- Escore – Extraversion
- Oscore – Openness
- Ascore – Agreeableness
- Cscore – Conscientiousness
- Impulsive – Impulsiveness
- SS – Sensation-Seeking

ROC Curves - Marijuana

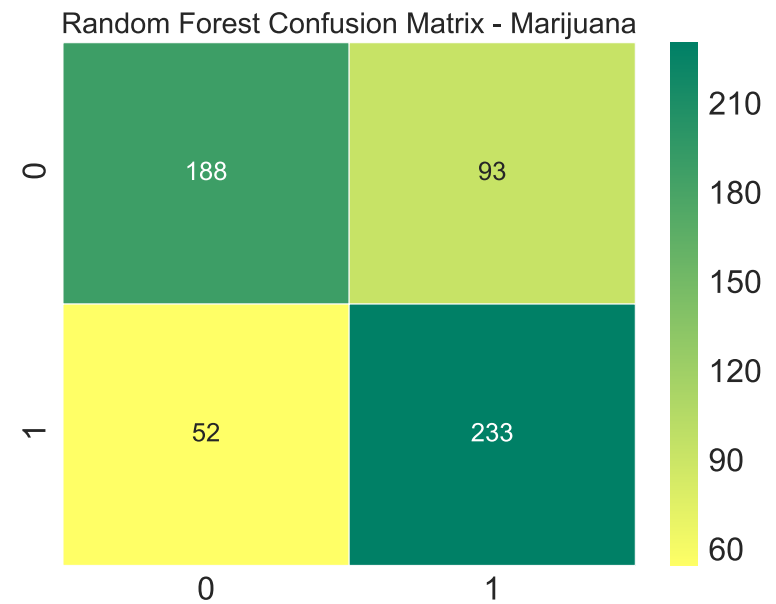
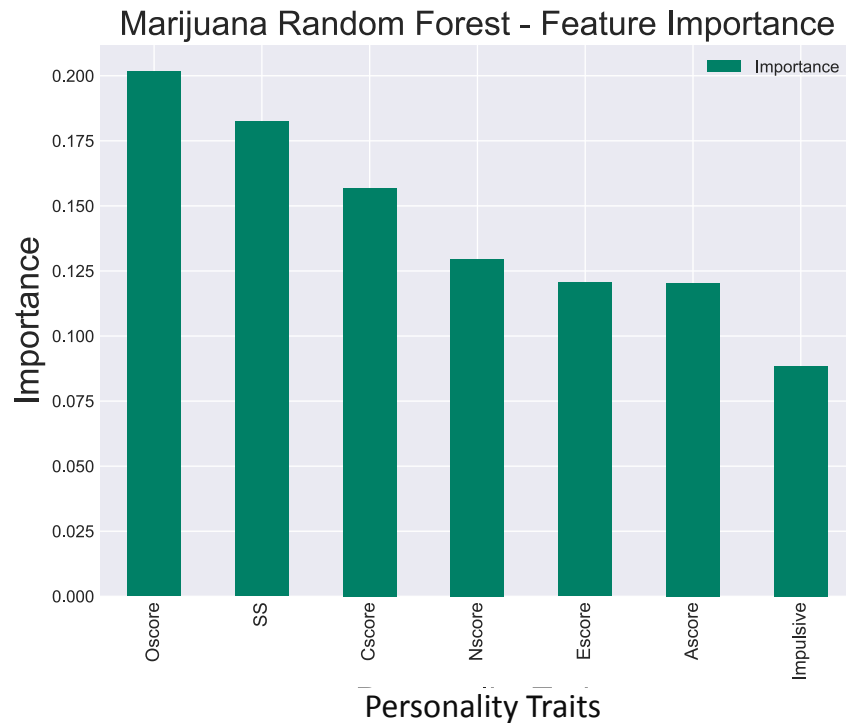


# Linear SVM



**Test Accuracy: 0.76% | Recall Class 0: 0.72 | Recall Class 1: 0.80**

# Random Forest



**Test Accuracy: 0.76% | Recall Class 0: 0.73 | Recall Class 1: 0.80**

# Flask

- Marijuana – Random Forest
- Nicotine – Linear SVM
- Cocaine – SMOTE, Linear SVM

<http://10.18.4.243:5000/predict>



# Findings - Marijuana

- Individuals that score higher on the openness and SS metrics are more likely to use marijuana
- Individuals that score higher on extraversion, conscientiousness and neuroticism are at less risk for using marijuana
- Agreeableness and impulsiveness are fairly neutral metrics

# Findings – Nicotine & Cocaine

- Individuals scoring higher on neuroticism and SS are at a greater risk for using both nicotine and cocaine
- High conscientiousness and extraversion scores tend to lead to less nicotine use
- High extraversion scores tend to lead to more cocaine use, while high agreeableness and conscientiousness lead to less cocaine use

**Thank You**

**Questions?**

# Appendix

# Multilabel Approach

- Random Forest – Multilabel Classification for 10 dependent variables

		precision	recall	f1-score	support
Heroin	0	0.93	1.00	0.96	525
	1	0.75	0.56	0.64	157
	2	0.83	0.89	0.86	288
	3	0.73	0.39	0.51	121
	4	0.73	0.56	0.64	147
Ketamine	5	0.25	0.03	0.06	32
	6	1.00	0.03	0.06	60
	7	0.68	0.56	0.61	108
	8	0.56	0.50	0.53	122
	9	0.73	0.80	0.76	318
avg / total		0.79	0.73	0.74	1878