

# Drunk and Disorderly Data: Applying a natural language processing algorithm to classify alcohol-related crimes in police data

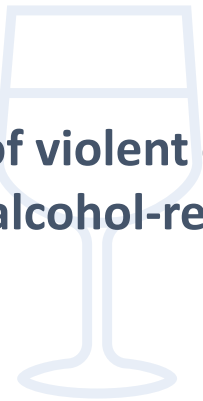
Olivia Horsefield

Mark Green

Carly Lightowlers

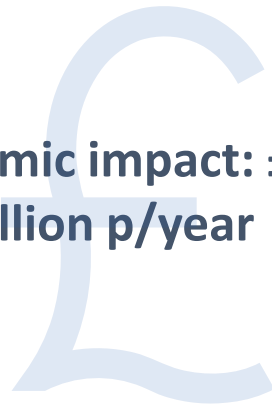
# Background | alcohol-related crime in the UK

**42% of violent crimes  
are alcohol-related**



Office for National  
Statistics (2021)

**Economic impact: £11  
billion p/year**



Home Office (2012)

**Social impact: involving  
injury, inequalities in  
victimisation**



Bryant (2020), Green et al  
(2017), Lightowlers et al  
(2021)

# Background | UK Data on alcohol-related crime

Secure data – police data

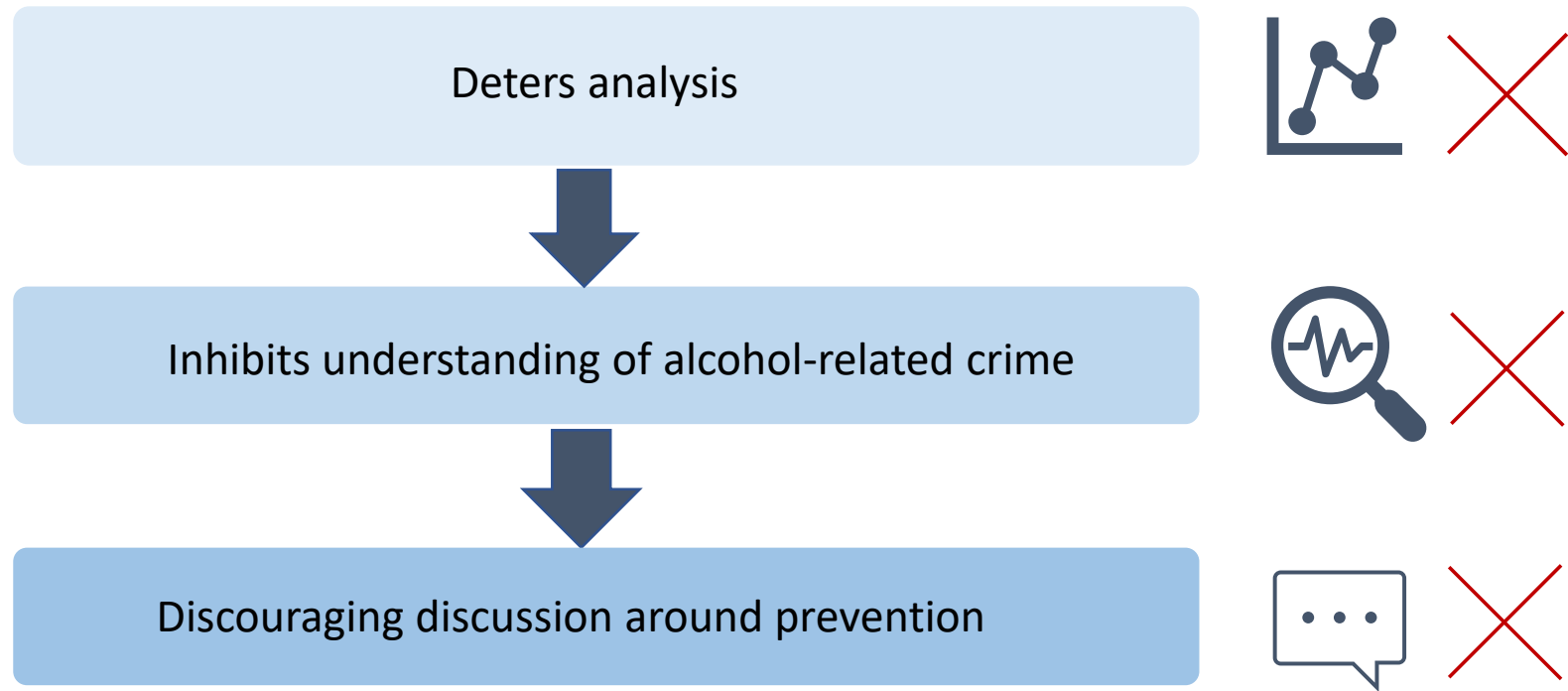
- Alcohol indicator
- Spatial
- Detailed – crime descriptions

# Background | Data on alcohol-related crime

Secure data – police data

- **Alcohol indicator**
- Spatial
- Detailed – crime descriptions

# Background | The consequences...



# Data

Merseyside Police

April 2021 – September 2022

All crime types captured

## Fields of interest

- Crime description/MO Field
- Alcohol indicator
- Time variables (month, day of week, time of day)
- Crime type



# Data

Merseyside Police

April 2021 – September 2022

All crime types captured

## Fields of interest

- **Crime description/MO** Field
- Alcohol indicator
- Time variables (month, day of week, time of day)
- Crime type

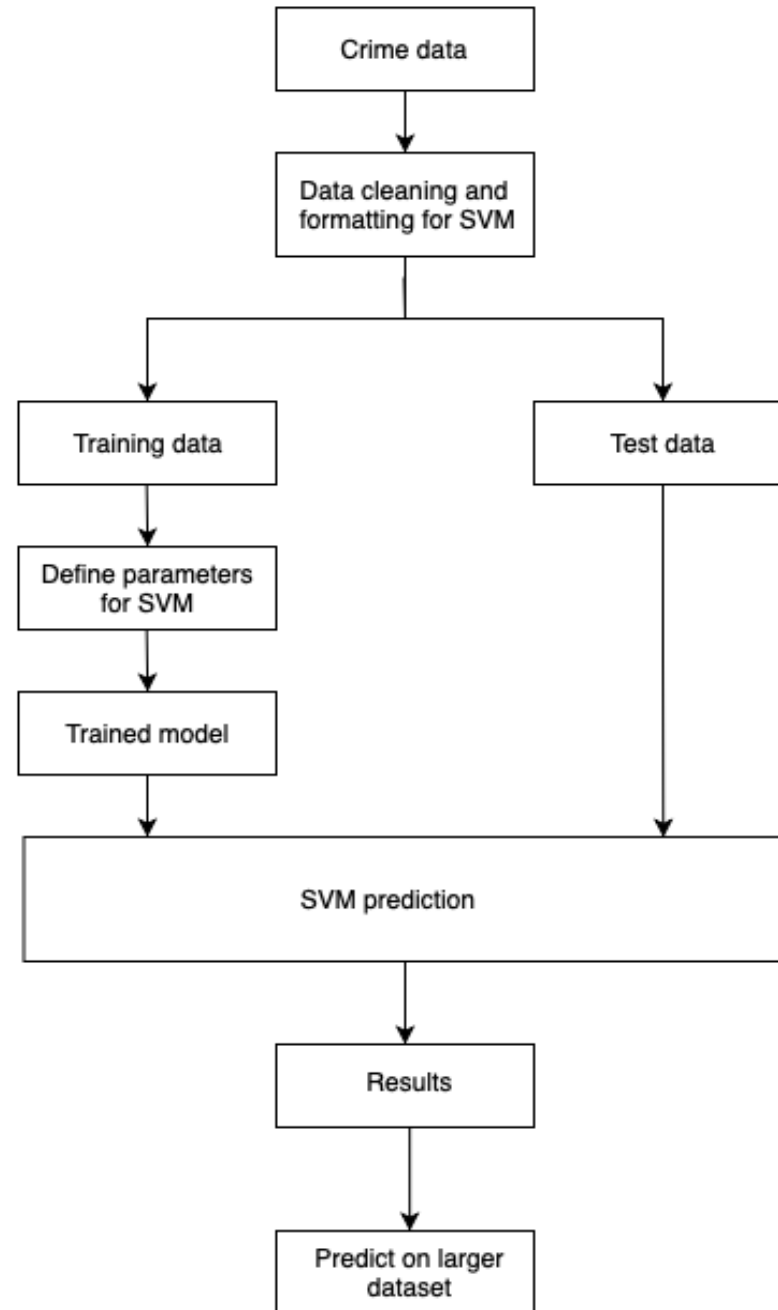


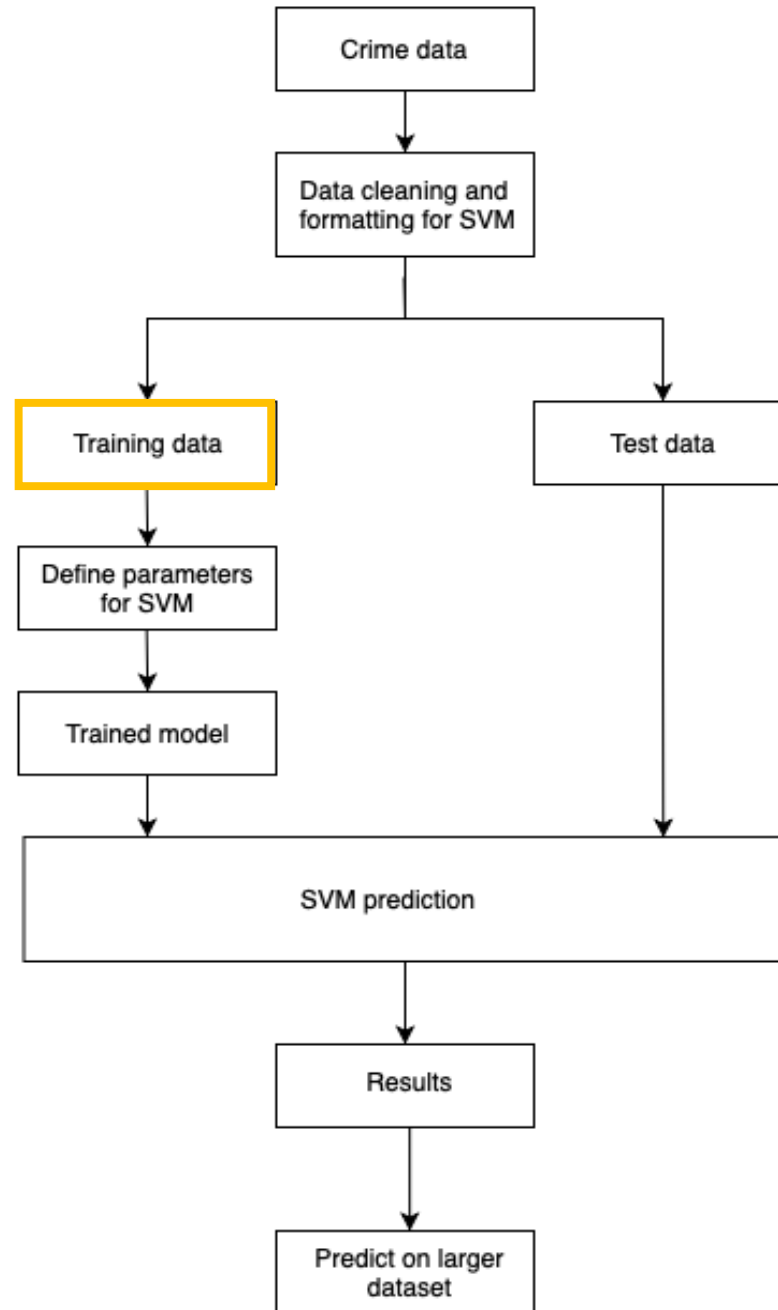
# Aim | A new measure – text classification

- Apply Natural Language Processing (NLP) Algorithm to crime descriptions and classify as alcohol-related
  - Decipher context
  - Applications in other areas of research



# Method | Support Vector Machine (SVM)





# Method | Support Vector Machine (SVM)

## Training (80% data)

1. Review crime descriptions and label with a 'true' alcohol label

Flag	Alcohol-related?
Crime description	✗
Crime description	✓
Crime description	✓
Crime description	✗

# Method | Support Vector Machine (SVM)

Training (80% data)

1. Review crime descriptions and label with a 'true' alcohol label

Flag	Alcohol-related?
Crime description	×
Crime description	✓
Crime description	✓
Crime description	×

# Method | Support Vector Machine (SVM)

## Training (80% data)

1. Review crime descriptions and label with a 'true' alcohol label

How do we determine if a crime is alcohol-related?

- Subjectively deciding
- Definitions
- Data analysis
  - 58% police flagged alcohol-related crimes contain violence-related terminology

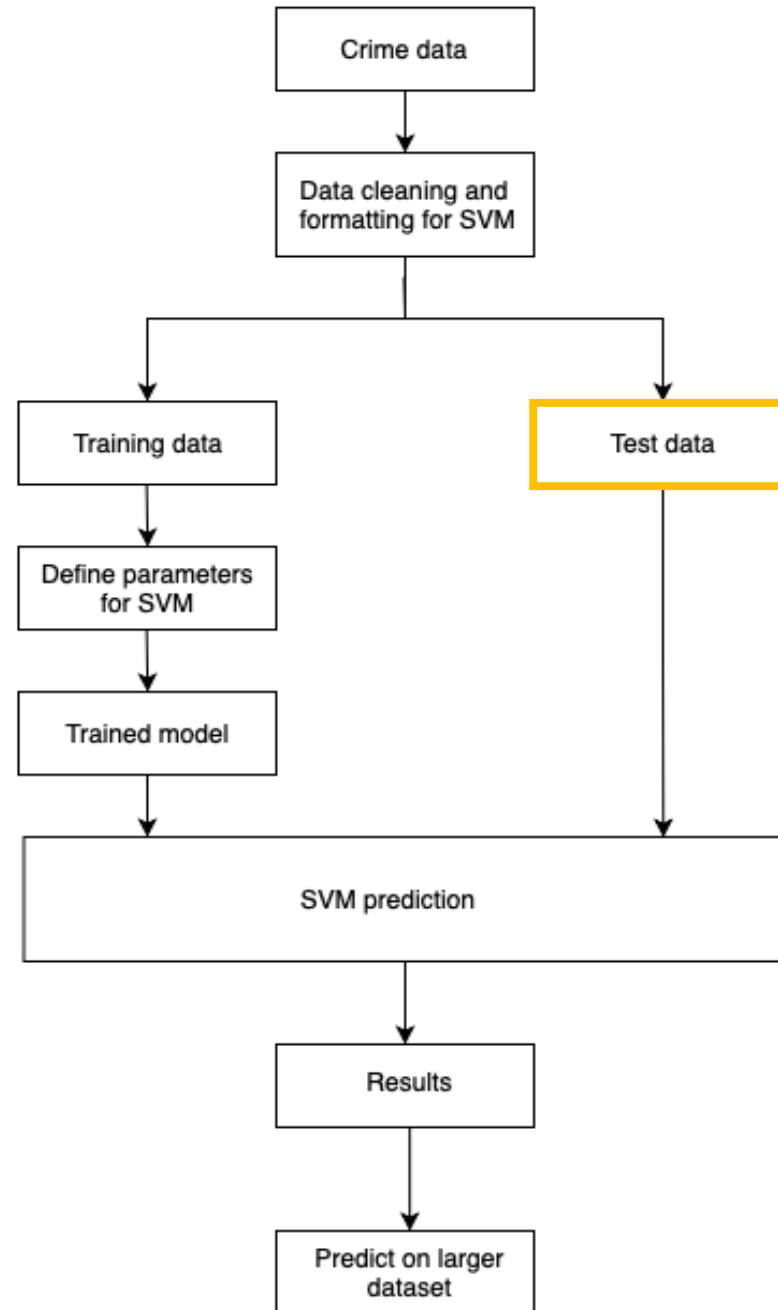
# Method | Support Vector Machine (SVM)

## Training (80% data)

1. Review crime descriptions and label with a 'true' alcohol label
2. Train model

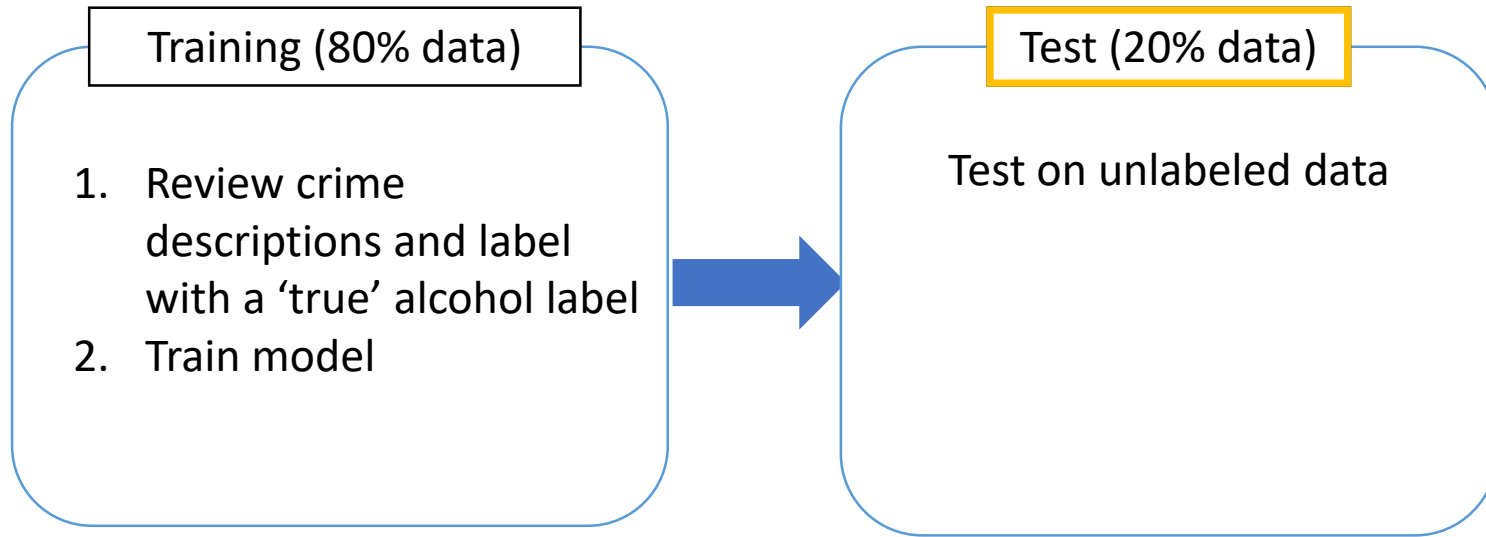
How do we determine if a crime is alcohol-related?

- Subjectively deciding
- Definitions
- Data analysis
  - 58% police flagged alcohol-related crimes contain violence-related terminology





# Method | Support Vector Machine (SVM)

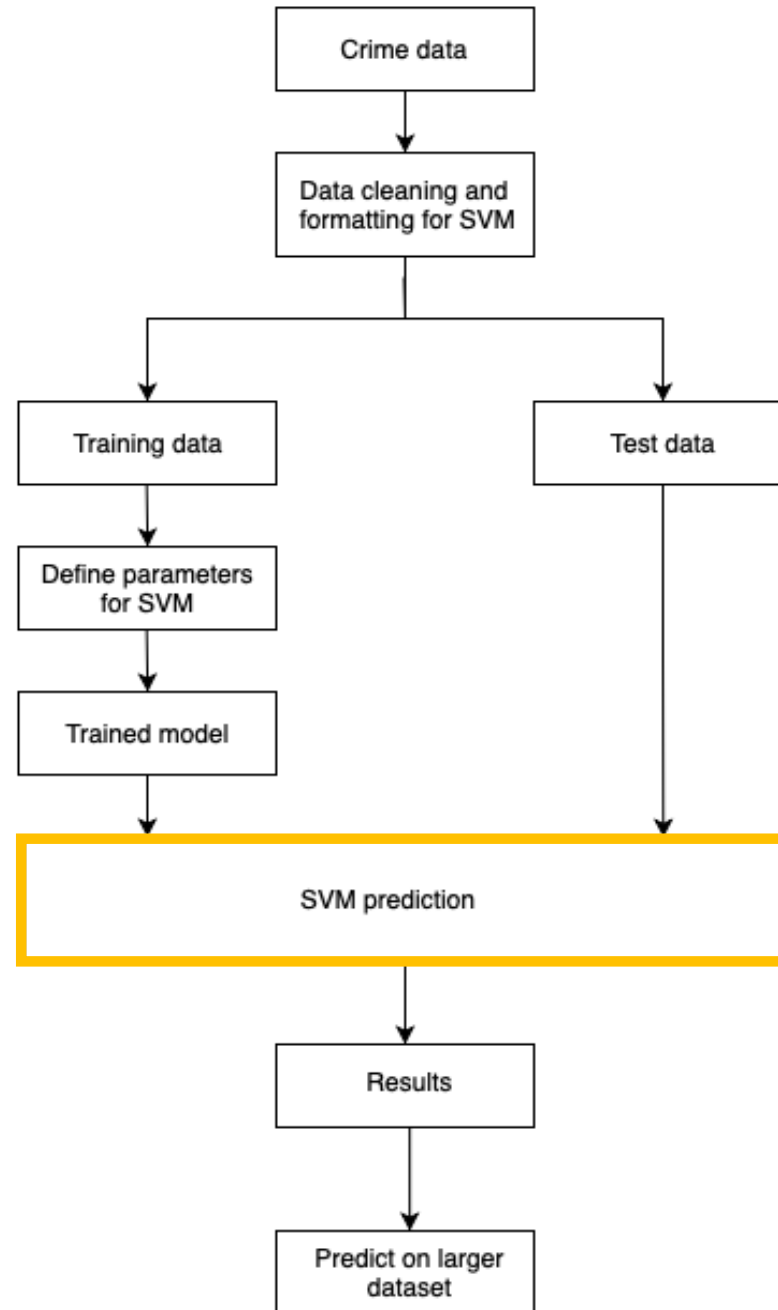


# Method | Support Vector Machine (SVM)

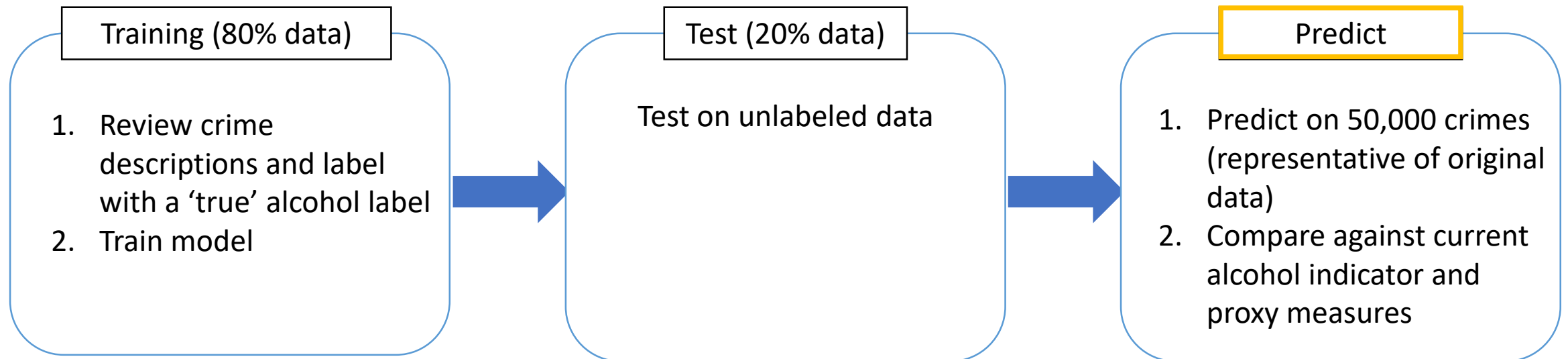
Test (20% data)

Test on unlabeled data

<b>Accuracy</b>	0.71
<b>Misclassification rate</b>	0.29
<b>95% CI</b>	0.62, 0.79
<b>Kappa</b>	0.42
<b>Sensitivity</b>	0.69
<b>Specificity</b>	0.73
<b>F-score</b>	0.71



# Method | Support Vector Machine (SVM)



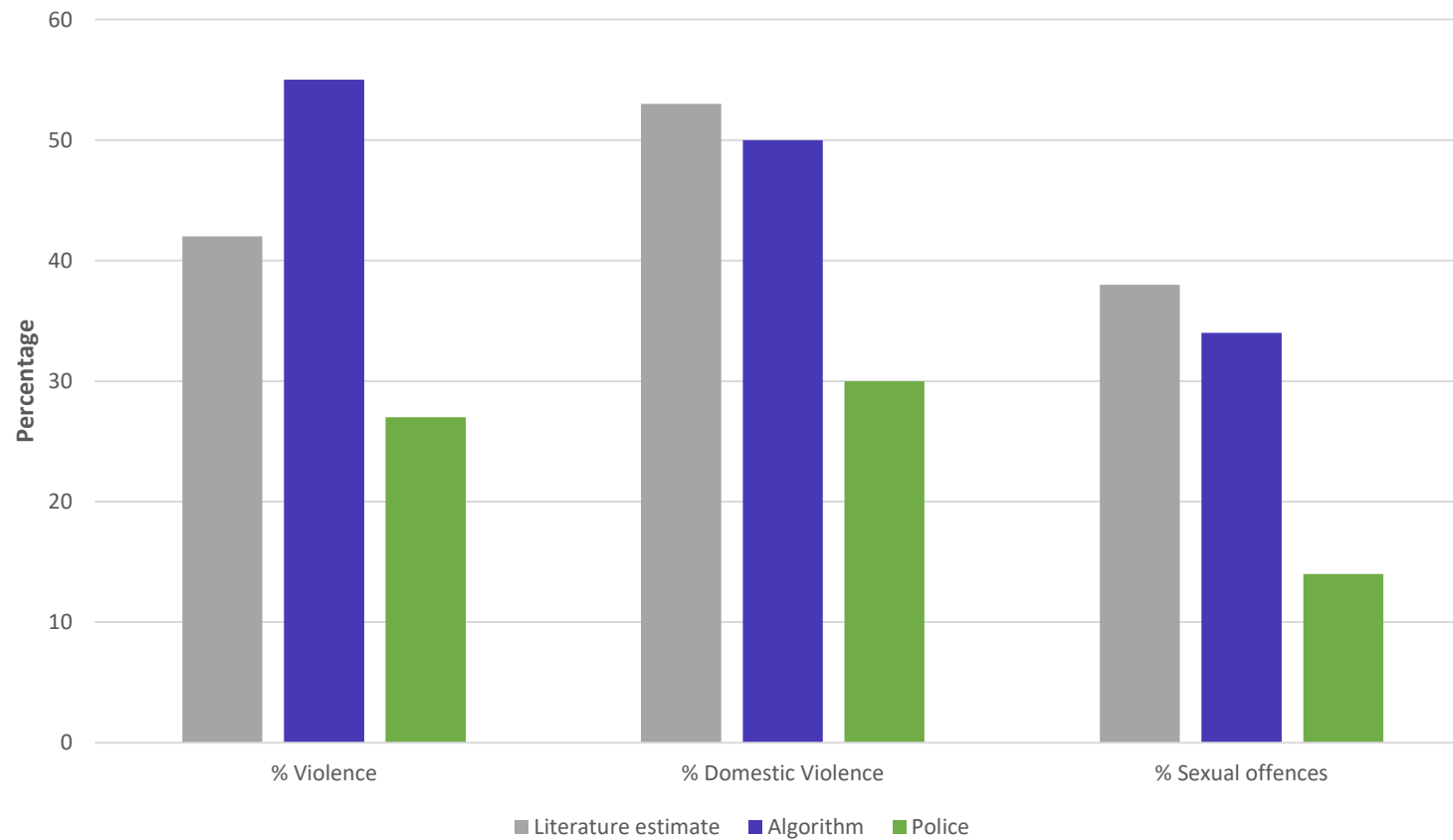
# Results

Flag	Count	Percentage
Police	6,398	12.8
Keyword	6,930	13.9
Time	8,234	16.5
Algorithm	15,526	31.1

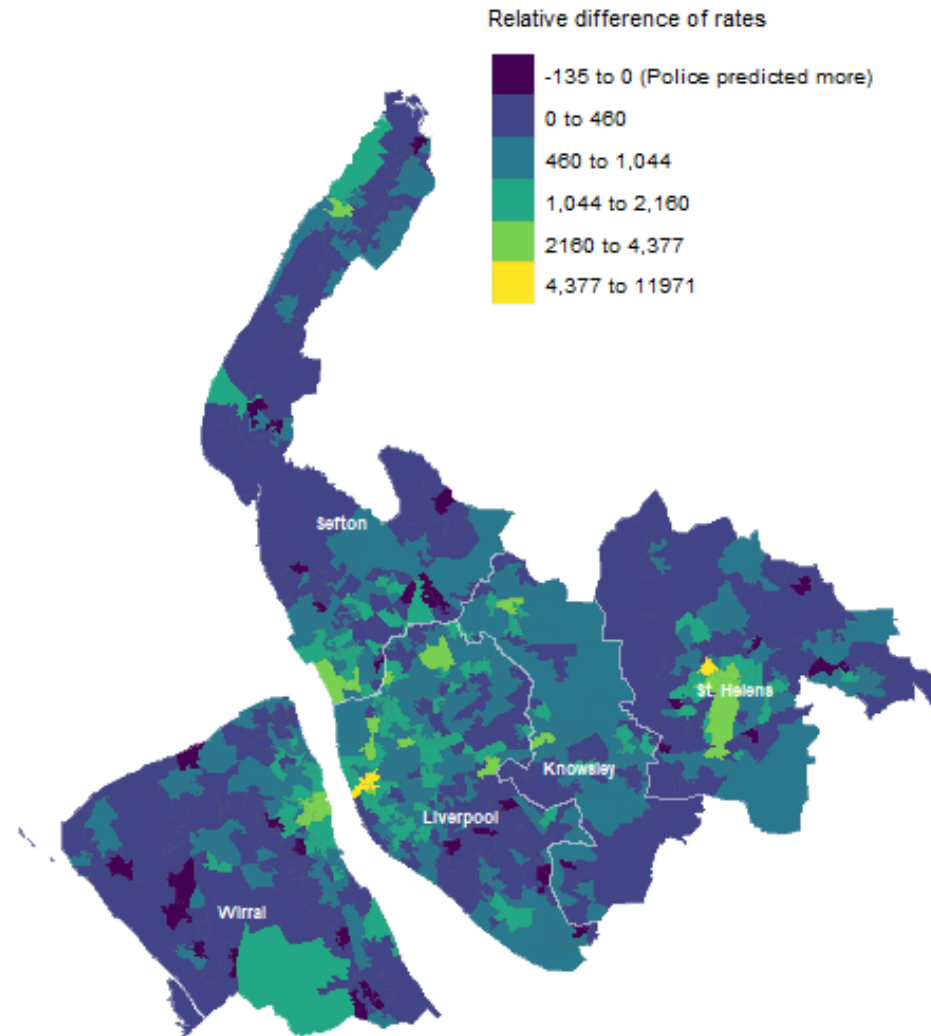
# Prediction validation – test accuracy on subset

Measure	Correctly predict alcohol-related crime
Police	76
Algorithm	82

# Results | Crime type



# Results | Alcohol-related crime mapped

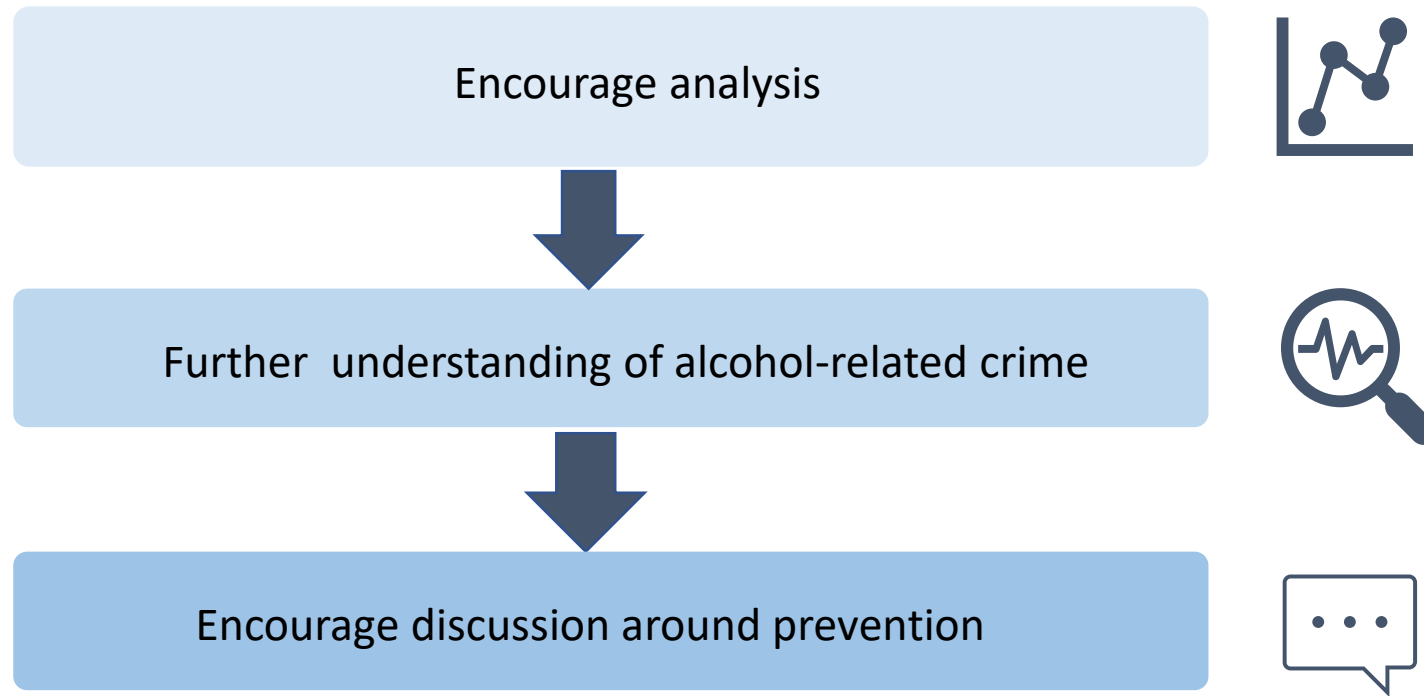




# Impact

Our approach may have improved *estimates* of alcohol-related crime,  
but also our understanding of its *nature*

# Impact



# Limitations

- Small training data size due to time restraints
- Ambiguity around definition
- Police data

# Drunk and Disorderly Data: Applying a natural language processing algorithm to classify alcohol-related crimes in police data

Thank you!



Olivia Horsefield