

ROWE

BY IDENTIFIERS



PSID: INTL-DA-07
TEAM LEADER NAME: VARTUL BAHUGUNA

////////

PROBLEM STATEMENT

Identification of crime prone area

...

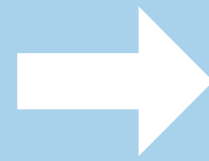




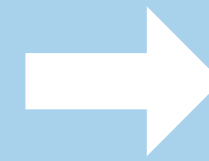
OUR METHODOLOGY



Data collection



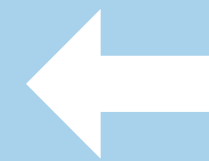
Pre-processing



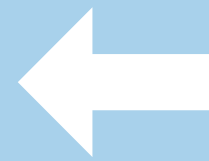
Model Selection



Model building



Verification of
accuracy



Visualisation

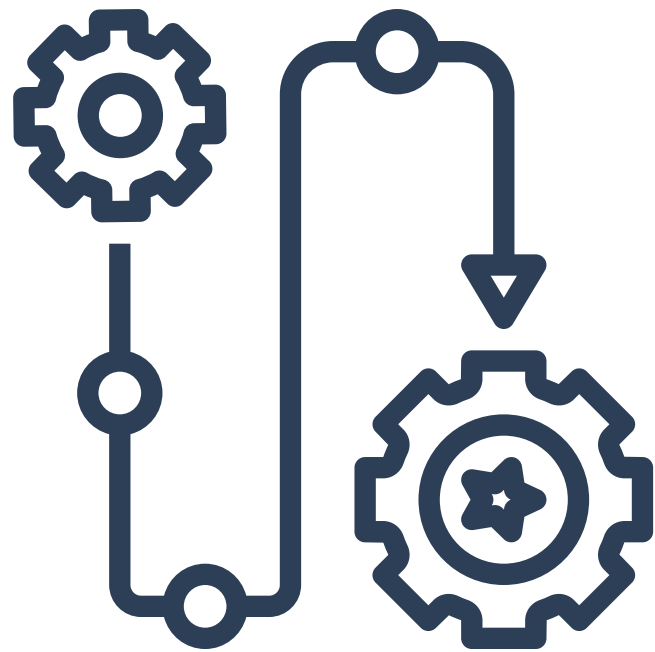


DATA COLLECTION AND AUGMENTATION

- Data set used is provided by Manthan's official site.
- Data is augmented for additional attributes; Gender and Age for more precise analysis.



PRE-PROCESSING



- TOKENIZATION: The tokens were extracted from the raw data provided.
- STOP WORDS: Non-essential English words used in a sentence are removed.
- LEMMATIZATION & STEMMING: were used to extract context of a word.



Event Type:
"Dispute between
rider and driver"



STOP WORD
REMOVAL



RESULT:
"DISPUTE"

PRE-PROCESSING PIPELINE



STOPWORDS REMOVAL AND TOKENIZATION

```
In [9]: df = df.dropna()                                #drop null values
df['Event Type'] = df['Event Type'].apply(str)          #converting floats and ints to string
ls = df['Event Type'].to_list()
n = len(ls)
ps = ls.copy()
for i in range(0,n):
    s = ls[i]
    stop_words = set(stopwords.words('english'))        #stop words removal
    word_tokens = word_tokenize(s)                     #tokenizing sentences
    filtered_sentence = [w for w in word_tokens if not w.lower() in stop_words]
    ls[i].join(filtered_sentence)
    ls[i] = ls[i].lower()
ls
```

```
Out[9]: ['information against police',
'threat in person',
'dispute',
'gambling',
'threat in person',
'missing',
'information against police',
'theft',
'dispute',
'dispute',
'dispute',
'domestic violence',
'threat in person',
'threat in person',
'threat in person',
'property disputes',
'information against police',
'property disputes',
'domestic violence',
```

FORMULATION OF INITIAL MODEL



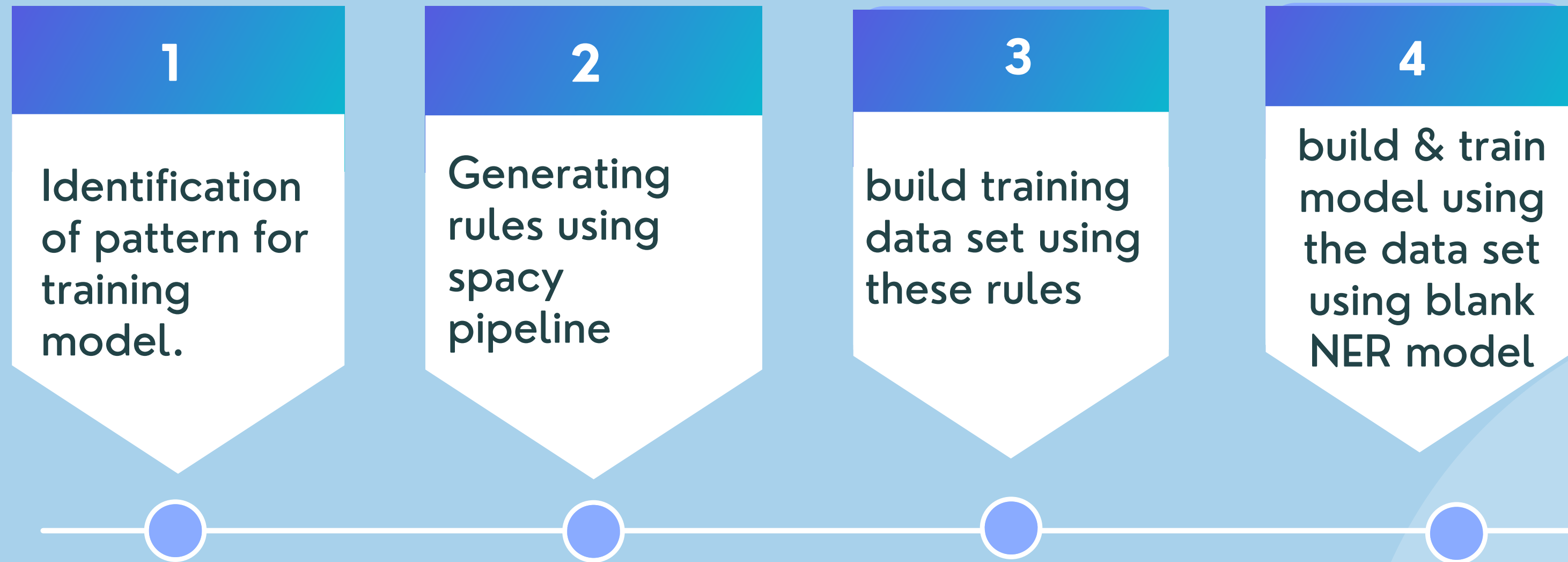
Identifying the recurring words that align with crime.

Classify obtained words in different levels on basis of severity.

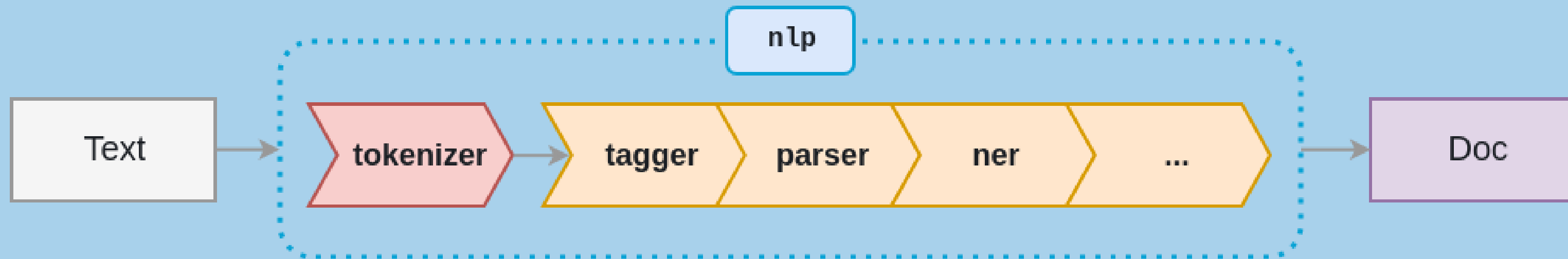
Classifying a crime in levels; if two different level of crimes are identified level assigned is one with higher order.



MODEL BUILDING



NER ARCHITECTURE



contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday's PaperAdvertisementSupported **ORG** byF.B.I. Agent Peter Strzok **PERSON** ,
Who Criticized Trump **PERSON** in Texts, Is FiredImagePeter Strzok, a top F.B.I. **GPE** counterintelligence agent who was taken off the special counsel
investigation after his disparaging texts about President Trump **PERSON** were uncovered, was fired. CreditT.J. Kirkpatrick **PERSON** for The New York
TimesBy Adam Goldman **ORG** and Michael S. SchmidtAug **PERSON** . 13 **CARDINAL** , 2018WASHINGTON **CARDINAL** — Peter Strzok
PERSON , the F.B.I. **GPE** senior counterintelligence agent who disparaged President Trump **PERSON** in inflammatory text messages and helped
oversee the Hillary Clinton **PERSON** email and Russia **GPE** investigations, has been fired for violating bureau policies, Mr. Strzok **PERSON** 's lawyer
said Monday **DATE** .Mr. Trump and his allies seized on the texts — exchanged during the 2016 **DATE** campaign with a former F.B.I. **GPE** lawyer,
Lisa Page — in **PERSON** assailing the Russia **GPE** investigation as an illegitimate “witch hunt.” Mr. Strzok **PERSON** , who rose over 20 years
DATE at the F.B.I. **GPE** to become one of its most experienced counterintelligence agents, was a key figure in the early months **DATE** of the
inquiry.Along with writing the texts, Mr. Strzok **PERSON** was accused of sending a highly sensitive search warrant to his personal email account.The
F.B.I. **GPE** had been under immense political pressure by Mr. Trump **PERSON** to dismiss Mr. Strzok **PERSON** , who was removed last summer
DATE from the staff of the special counsel, Robert S. Mueller III **PERSON** . The president has repeatedly denounced Mr. Strzok **PERSON** in posts on

Name Entity
Recognition helps in
detecting a named
entity and categories it.

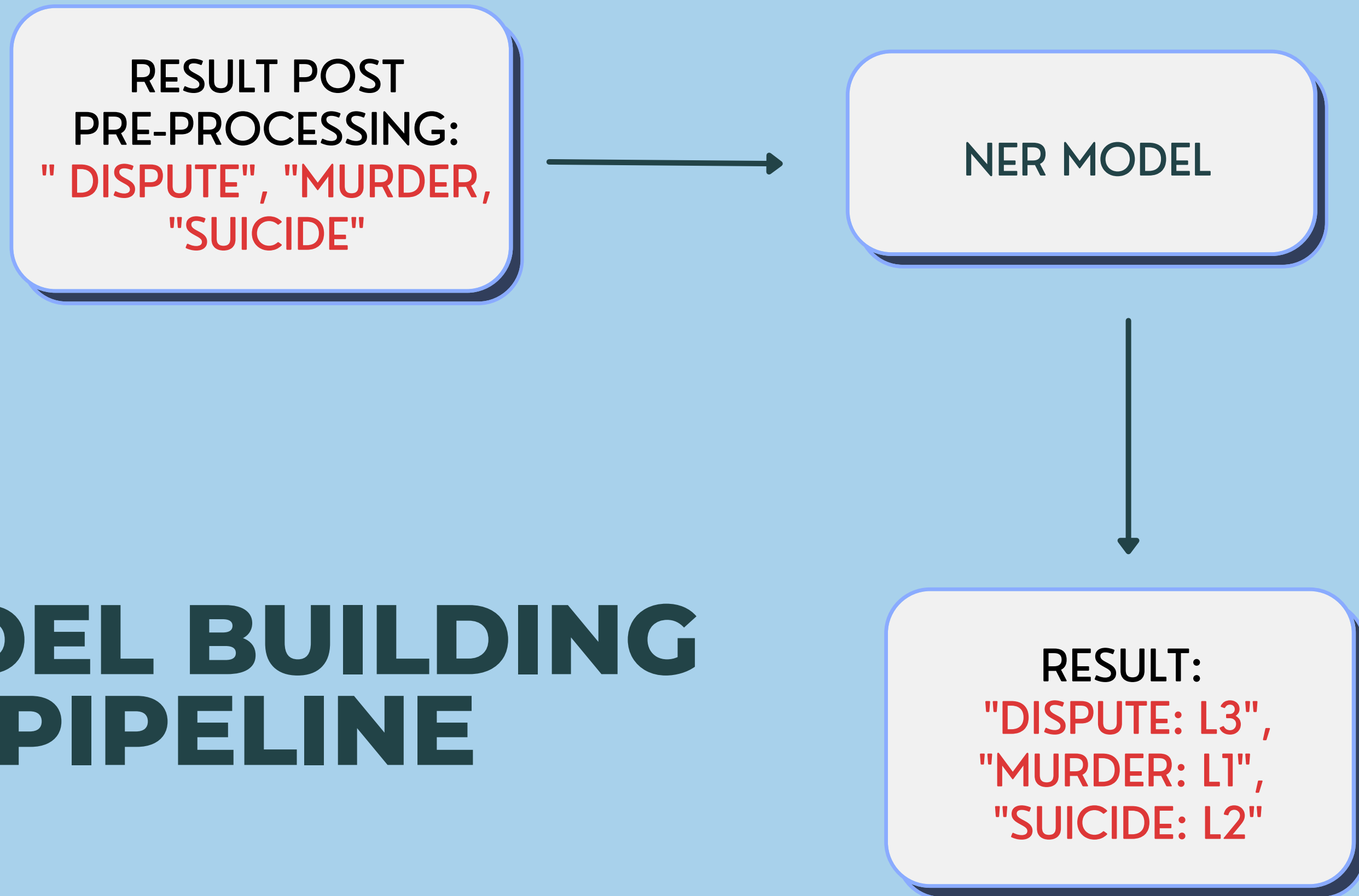


NER MODEL IMPLEMENTATION USING SPACY LIBRARY

Spacy is an open-source software python library used in advanced natural language processing and machine learning. It will be used to build information extraction, natural language understanding systems, and to pre-process text for deep learning.

As spaCy uses the latest and best algorithms, its performance is usually good as compared to NLTK

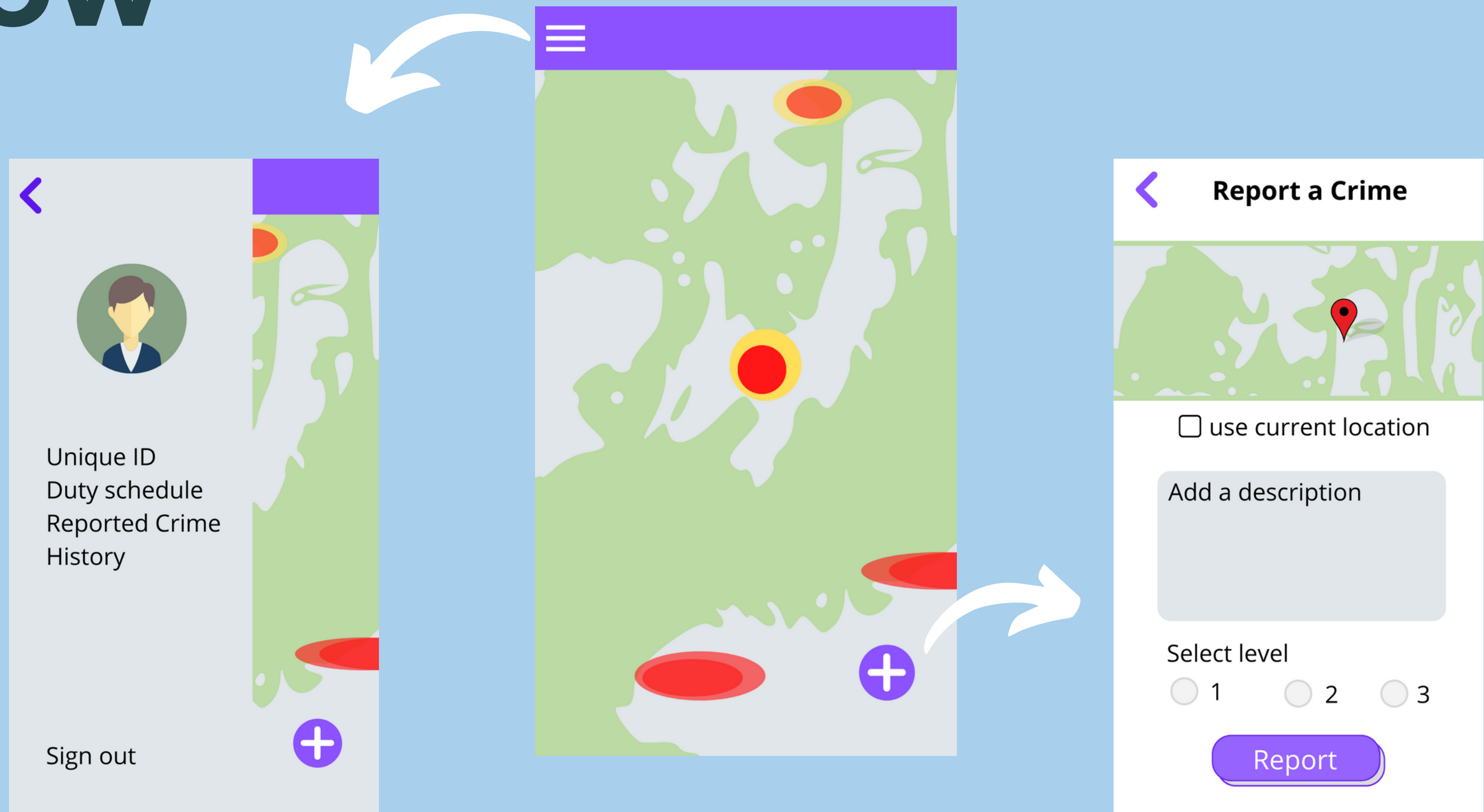




MODEL BUILDING PIPELINE



PROPOSED UI FLOW



THANK YOU

