



—
INSTITUT DE HAUTES
ÉTUDES INTERNATIONALES
ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE
OF INTERNATIONAL AND
DEVELOPMENT STUDIES

How may the contestation of the definition of national identity by different political camps relate to their attitudes towards asylum seekers? Using XGBoost to analyse the newspaper articles in Hong Kong

THESIS

submitted at the Graduate Institute
in fulfillment of the requirements of the
The International Relations/Political Science Department

by

Chin Man Kwan

Thesis No.

Supervisor: Prof. Sungmin Rho
Second Reader: Prof. Ravinder Bhavnani

Geneva
2022

RESUME / ABSTRACT

Asylum seekers (or formally non-refoulement claimants) have started to draw attention in Hong Kong's public debate since early 2016 after a senior official from the Immigration Department claimed that the city's mechanism of screening asylum seekers was being abused. Not unseen in other places, this group of vulnerable population has then started to be portrayed by some media outlets pejoratively, such as being questioned for the genuineness of their non-refoulement applications as "fake refugees". When Hong Kong embroiled into one of its largest conflict between pro-democracy protesters and the HKSAR Government in 2019 and both the pro-Beijing and opposition camps manifested their respective political ideology saliently, how might the depiction of non-refoulement claimants vary according to the political camp that media outlets are linked to? This thesis aims to investigate this question by using machine-learning-based sentiment analysis on news articles reporting on non-refoulement claimants published in 2019 by printed newspapers in Hong Kong. It is found that, in general, pro-Beijing camp media outlets were more likely to portray non-refoulement claimants negatively than outlets with other political stances. With the promulgation of the National Security Law in mid-2020 and the recent closures of several pro-democracy newspaper outlets, one can expect that the narrative on asylum seekers in Hong Kong will be further dominated by pro-Beijing media outlets, thereby potentially fostering a more hostile social environment against those who seek refuge in Hong Kong due to facing threats of tortures in their home countries.

Acknowledgements

Throughout the writing of my thesis and my study at the Graduate Institute, I am very grateful to have received an enormous amount of support and assistance.

Starting with my thesis, I would like to thank my supervisor, Professor Sungmin Rho, who has provided invaluable and timely feedback on how I should proceed with my research, especially when at times I was stuck at how to proceed. Also my deepest gratitude to Professor Ravinder Bhavnani for agreeing to be my second reader.

I would also like to thank all the professors and teaching assistants at the IRPS department for all of the amazing courses they have taught. I have surely come across a lot of stimulating and intriguing topics and perspectives throughout the courses I have taken. Particularly, I had never expected that I would discover my interest in data science via the two stats courses offered by the department.

No matter it is emotional support during stressful times or academic support towards each other on a variety of subjects, it would have been a much more difficult academic journey for me without the support from my cohort at the IRPS programme. Hope we can all hang out together one day.

I also express my deepest gratitude towards the developers of all the packages I used in this thesis. Without them, I would not have been able to complete my analysis. They will always be indispensable for my data science toolbox. Also a big thank you to the developers of the `iheiddown` package which helped enormously in formatting my thesis.

On a personal level, my beloved parents have offered their unlimited support to me throughout my study at the Institute and my life. Without them I would not have been able to be who I am today. My friends back in Hong Kong whom I have known since high school have never ceased to be an amazing companionship even if we are now living in two different continents. I have also known some of the most agreeable friends at the Institute, and they have always cheered me up in times of difficulty.

Lastly, to whoever is reading this thesis, thank you for spending your time on doing so. I hope it will be worth your time, and you will enjoy it.

Chin Man KWAN
Graduate Institute, Geneva
10 January 2022

Contents

Preface	1
1 Introduction	4
1.1 Contribution to the literature	6
1.2 Outline	8
2 Literature Review	9
2.1 Theories	9
2.1.1 What is national identity?	9
2.1.2 Impact of conceptions of national identity on attitudes towards immigrants	11
2.1.3 Chineseness in Hong Kong	13
2.2 Empirics	15
2.2.1 National identity of Hong Kong defined by different political camps	15
The pro-Beijing camp	16
The pro-Democracy camp (or pan-democrats)	18
The localist camp	20
2.2.2 Hong Kong's non-refoulement policy regime	23
2.2.3 How do different political camps respond to the issue of asylum seekers in Hong Kong?	26
Theoretical expectations in the case of Hong Kong	26
Empirical observations	27
3 Methods	30
3.1 Hypothesis	30
3.2 Data collection	31
3.3 Machine learning analysis	34
3.3.1 Model selection	34
3.3.2 Data pre-processing	35
3.3.3 Implementation	36

CONTENTS

3.4	Operationalisation	37
3.4.1	Main independent variable (a.k.a main feature)	37
	Publication of news articles by outlet and political camp	38
3.4.2	Dependent variable (a.k.a target variable)	40
3.4.3	Control variables (a.k.a features)	42
4	Results	43
4.1	Exploratory data analysis (EDA)	43
4.1.1	Number of news articles by month	43
4.1.2	Polarities of the news articles	43
4.1.3	Presence of racial labels	45
4.1.4	Character lengths of news articles and titles	45
4.2	Sentiment analysis	48
4.2.1	Pre-processing	48
4.2.2	Training the model	51
4.3	Results: was pro-Beijing media outlets more likely to negatively portray asylum seekers in Hong Kong in 2019?	53
4.3.1	Overall magnitudes of SHAP values for each feature	53
4.3.2	SHAP values of predicting negative news articles	55
4.3.3	SHAP values of predicting neutral news articles	58
4.3.4	SHAP values of predicting positive news articles	61
4.3.5	Problem and solution: a topic with multiple angles by different camps?	64
4.4	Discussion	65
4.4.1	Snapshots of news articles	68
4.4.2	Validity of automated sentiment analysis	70
5	Conclusion	71
5.1	Limitations	71
5.2	Future research directions	71
5.3	How might the instigation of the National Security Law affect the public discourse on asylum seekers in Hong Kong?	72
	Appendix: The Echoes of the Code	74
	Appendix: The Echoes of the Code redux	75
	References	78

List of Figures

2.1	Number of non-refoulement claim cases (left) and refugees (right) in Hong Kong .	24
2.2	Number of non-refoulement claimants by nationality in 2021	25
3.1	News articles on asylum seekers in 2019 by news outlet (left) and political camp (right)	39
4.1	Temporal patterns of the publication of news articles about asylum seekers in Hong Kong in 2019	44
4.2	Presence of racial labels in the news articles by sentiment	46
4.3	Distributions of the word counts of the articles' titles (left) and main texts (right)	47
4.4	Elbow plot of the reconstruction error of NMF as a function of the number of pre-specified latent topics	50
4.5	Word list of each of the topics generated by the NMF model, note: the importance of each word in each topic is arranged in descending order from left to right	51
4.6	Magnitudes of SHAP values of each feature for all classes	54
4.7	The SHAP values of the features in the prediction of whether an article has a negative polarity	56
4.8	Dependence plot of the eight most important features for predicting negative polarity of news articles	57
4.9	Dependence plot of the eight most important features for predicting negative polarity of news articles	57
4.10	The SHAP values of the features in the prediction of whether an article has a neutral polarity	59
4.11	Dependence plot of the eight most important features for predicting neutral polarity of news articles	60
4.12	Dependence plot of the eight most important features for predicting neutral polarity of news articles	60
4.13	The SHAP values of the features in the prediction of whether an article has a positive polarity	61

LIST OF FIGURES

4.14 Dependence plot of the eight most important features for predicting positive polarity of news articles	62
4.15 Dependence plot of the eight most important features for predicting positive polarity of news articles	63
4.16 Interaction between pro-Beijing affiliation of newspaper outlets and the salience of the problem and solution topic for the XGBoost model's prediction of the negative (left) and positive (right) polarities	66
4.17 Excerpt from pro-Beijing Oriental Daily News	68
4.18 Excerpt from pro-Beijing Ta Kung Pao	69
4.19 Excerpt by pro-Democracy Apple Daily	69

List of Tables

2.1	Comparison of the three camps on key dimensions of Hong-Kong-China relations and their attitudes towards non-refoulement claimants (the last row)	22
3.1	Translation of the keywords used to search for media reports on WiseNews	33
3.2	Classification of the associated political camp of media outlets in Hong Kong included into the sample, source: Fong (2017a); Lee (2018); author's own research	37
4.1	Polarities of the news articles on asylum seekers in Hong Kong in 2019	45
4.2	Summary statistics of the word counts of the news articles' titles and main texts	46
4.3	F1 scores on 5-fold cross validation and test data of the baseline models	52
4.4	F1 scores on 5-fold cross validation and test data of the baseline and tuned XGBoost models	52
4.5	F1 score breakdown of the tuned XGBoost model on the test data	53

Preface

My motivation to write for the topic of media portrayal of asylum seekers (or more formally, non-refoulement claimants) in Hong Kong mainly comes from two sources. On one hand, the salient political affiliation of media outlets in Hong Kong as recorded by Lee (2018) has greatly affected how various socio-political issues are framed in the public debate, especially during the 2019 anti-extradition law protest which mobilised millions of Hong Kong citizens to first oppose a law that would have allowed fugitives to be extradited to mainland China and then fight for democracy in the city's political system. As much as these high-profile issues directly relating to the political authority and control over the city by the Beijing Government undoubtedly mark the distinction between pro-Beijing and non-pro-Beijing media outlets in terms of reportage, I suspect that the difference in the narratives by the media outlets in Hong Kong about non-refoulement claimants can also be demarcated based on whether a media outlet is affiliated with the pro-Beijing camp or not.

Indeed, the Hong Kong Government adopts an unwelcoming stance towards foreigners who wish to seek asylum within the city's territory. Apart from completely rejecting the possibility of any refugees being recognised and resettled locally (Immigration Department 2021b), the Hong Kong Government also becomes more suspicious about the genuineness of non-refoulement claims in recent years as the former Chief Executive *CY Leung* once suggested that Hong Kong leave the UN Convention against Torture in 2016 to alleviate the so-called "fake refugee" problem (Cheng 2016b). Therefore, pro-Beijing media outlets may echo the Government's unwelcoming stance towards non-refoulement claimants in their reportage to sway the public opinion in favour of more stringent measures, whereas pro-democracy outlets may offer counter-narrative which are more sympathetic towards this group of vulnerable population.

LIST OF TABLES

The increasing political pressure faced by alternative media in Hong Kong and, by the beginning of 2022, the closures of several prominent pro-democracy media outlets which had served to be a crucial alternative sources of information about socio-political issues for the public besides foreign media outlets, further increases the rationale for studying whether the narratives on non-refoulement claimants between pro-Beijing and non-pro-Beijing media outlets may differ in terms of attitudes. This is because if political affiliation does turn out to be an influential factor of predicting the sentiment of news articles about asylum seekers in Hong Kong and pro-Beijing camp newspapers are indeed more negative in their coverage about non-refoulement claimants, then it is likely that the media's narrative on asylum seekers in the city will be increasingly dominated by pro-Beijing media outlets following the dissolution of pro-democracy media outlets. This may then foster a higher degree of hostility against non-refoulement claimants on top of the Hong Kong Government's already unwelcoming policies.

Apart from my interest in the media landscape of Hong Kong, I also took this opportunity to test whether automated approaches can help effectively learn the relationship between the media outlet's political affiliation plus other features derived from the news articles and their sentiments towards asylum seekers. As van Atteveldt et al. (2021) note, automated sentiment analysis for non-English language sources appears to be rarely used and yet, despite some caveats, this method can potentially cope with problems of scaling research on larger dataset faced by manual coding. Given the prominence of text as a major medium of transmitting information to the readers in newspapers and the availability of tools and software specifically designed for natural language processing in Chinese, studying whether political affiliations of newspaper outlets may be correlated to the attitudes of the news articles on non-refoulement claimants via machine-learning-based sentiment analysis can help explore whether such automated approaches can also generate reasonable insights from Chinese language text data for social sciences research. If machine learning models perform adequately in learning about the relationships between different features (some of which are created from the text in the articles) and the sentiment of the news articles in this analysis, then this may imply that future social sciences research on textual data (e.g. sentiment analysis) in Chinese language can also benefit from using automated methods to scale on large dataset, given that nowadays a huge amount of information in textual format is generated each day which will be difficult for manually going through all the entries in a dataset.

LIST OF TABLES

The process of the research for this thesis was quite long and not without difficulties. First of all, given the vast volume of literature related to the study of asylum seekers based on numerous perspectives and levels of analysis, it was quite easy to often get lost while reviewing the literature. In the end, I decided to focus on the literature of how the definition of national identity may affect the perceptions towards immigrants (of which asylum seekers are a subgroup) because of the remarkable ethnic difference between most of the ethnic-Chinese citizens in Hong Kong and non-refoulement claimants. Besides, the political division in Hong Kong is arguably more salient on how should the national identity of citizens in Hong Kong be defined (e.g. whether ethnic Chineseness or civic values should play a larger role in the definition) even though it is equally valid to classify the political parties within each camp according to the more conventional left-right socio-economic dimensions (Ma 2012).

The data collection and preprocessing part also took quite a considerable amount of time. Apart from filtering out relevant news articles about non-refoulement claimants published by newspapers in 2019, the unstructured nature of text data also required a lot of steps to transform them into a suitable format for being machine learning model inputs. Moreover, the fact that the downloaded news articles from *Wiseneews* database do not come with pre-labelled sentiments of the articles also means that it is important to refer to reliable coding scheme to operationalise the dependent variable of this thesis (fortunately the literature is always here to help). As much as there are numerous performant machine learning models available nowadays, the results will not be very meaningful if the data quality is not optimal. Modelling the data, on the other hand, was a relatively less difficult process than the data preprocessing part.

In any case, this thesis was finally written. This also marks the end of my master's programme at the Graduate Institute. Overall, this was quite an enriching experience. I can't simply describe with words how much I have learnt over the last two years. I believe this experience will be invaluable for my life.

Chapter 1

Introduction

Asylum seekers, defined by *Refugees, Asylum-seekers and Migrants* (2022) as individuals who are forced to flee from their home countries due to threats of persecutions and human rights violations, has become one of the most salient socio-political issues in the 21st Century, especially after the 2015 European Refugee Crisis which has then led to numerous political consequences, notably the rise of anti-immigration parties in several European countries. Just as the European Refugee Crisis started to unfold in 2015, the issue of asylum seekers (or more formally, non-refoulement claimants) had not gain significant attention in Hong Kong until early 2016, after a senior Immigration Department official claimed that the city's Unified Screening Mechanism (USM) used for assessing non-refoulement applications was being abused during a TV interview and two fabricated WhatsApp audios alleging asylum seekers of committing robberies and homicides in the New Territories were propagated (Ng, Choi, et al. 2019). Not unseen in other countries, non-refoulement claimants are often depicted with pejorative terms questioning the genuineness of their applications to seek refuge in Hong Kong from threats of tortures back in their home countries. For instance, during the months when non-refoulement claimants first attracted attention in public debate, a large proportion of headline articles in newspaper were published by two pro-Government (or pro-Beijing) newspapers which portrayed this group of population with various derogatory terms, such as being “fake refugees” who exploit the city's USM to seek illegal work or commit crimes (Ng, Choi, et al. 2019).

The broader political context of Hong Kong as a destination of asylum seekers is that since the transferral of sovereignty from Britain to China in 1997, the city has witnessed increasing tension with its new sovereign owner. Until July 2020 when the *Hong Kong Special Administrative Region*

1. Introduction

(HKSAR) Government passed the National Security Law and the subsequent commencement of the large-scale suppression of the opposition camp such as the mass arrest of 53 pro-democracy activists in January 2021 (BBC 2021), protests against the authority of the HKSAR and Beijing Governments had often occurred. Apart from the annual July 1 protest since 2003, the two most prominent political movements would be the 2014 Umbrella Movement and the more recent 2019 anti-extradition law protest which created considerable attention from other countries as well. In the latter protest, what started from several marches in June that altogether mobilised millions of Hongkongers to protests against a law which would have allowed fugitives in Hong Kong for being extradited to mainland China eventually evolved into months-long protests including several violent clashes between protesters and the Hong Kong police. Indeed, the political identities of both pro-Government and opposition camps manifested saliently and confronted each other during the 2019 protest. Perhaps one of the most high-profile events of the confrontation between the two camps' identities would be some protesters forcing their way into the central chamber of the city's legislature building while vandalising the HKSAR emblem and raising the former British colonial flag on the very day which was the 22nd anniversary of Hong Kong's sovereignty being transferred to China (BBC 2019).

The question that this thesis will investigate with regard to these two events is:

Given the highly politicised nature of Hong Kong's mass media landscape as recorded by Lee (2018) and the prominent role of the Chinese-language mass media in Hong Kong in framing socio-political issues, is the sentiment of news articles towards asylum seekers in Hong Kong correlated with the political affiliation of newspaper outlets?

The data to be used for investigating this question will be Chinese-language newspaper articles published in 2019 on non-refoulement claimants. The reasons for focusing on news articles from 2019 are twofold. Firstly, I would like to investigate if the publication pattern (e.g., the dominant publishing outlet on this issue and the distribution of the sentiments of the articles) of news related to asylum seekers in Hong Kong in 2019 shared any similarities when compared to the issue first became salient back in 2016. Secondly, since 2019 was marked with one of the most intense conflicts between the Government and oppositions in post-1997 Hong Kong, it would also be interesting to see if there exists pronounced difference in how non-refoulement claimants might

1. Introduction

be portrayed by media outlets linked to different political camps, just as how the pro-Beijing and opposition camps confronted each other on the streets.

1.1 Contribution to the literature

Besides the motivation to write on this topic as mentioned in the *preface* section, this thesis also aims to make the following contributions to the literature. Starting with the literature on asylum seekers in Hong Kong, although there certainly exist numerous studies on non-refoulement claimants with various perspectives such as an anthropological approach to study their life in the city (Vecchio 2014), how asylum seekers may be served for the identity construction of some youngsters (Mathews 2018) and how Hong Kong's policy and legal apparatus marginalise this group of vulnerable population (Ng 2020), to date there are only few studies directly related to how asylum seekers are portrayed by mass media shortly before and after the issue first drew public attention in 2016 which was an election year of the city's legislature (Ng, Choi, et al. 2019; Ng, Choi, et al. 2021). While these two studies (one on traditional and another on online news media) offer invaluable inspiration to this thesis and are important precursors for the study on this topic, it would be beneficial to also study how news articles about asylum seekers are covered some time after the issue first gained public attention to see if the publication pattern remains consistent or not. This can help better understand if media outlets from certain political camps were likely just galvanising the issue of non-refoulement claimants for electoral gains of their camp in the short run, or they are also attempting to construct a particular framing and narrative on asylum seekers within the city which may then be used to justify policy-making positions in the long run.

This thesis also aims to contribute to the literature on the effects of different definitions of national identity on the perception of immigrants. Specifically, I am interested in whether definitions of national identities proposed by different political camps which vary their emphases on the ethnic-based or civic-based dimensions (the definitions of these dimensions will be elaborated in the *Literature Review* chapter) may lead to different attitudes towards asylum seekers as reflected in the media discourse produced by each camp. It is certainly true that Hong Kong has its own characteristics which may not be applicable in other cases (for example, Hong Kong has never been a signatory of the 1951 Refugee Convention and thus is not bound to recognise refugees),

1. Introduction

but on a more abstract level, the definitions of Hong Kong's national identity by different political camps can be categorised according to the ethnic and civic dimensions (as discussed in next chapter), and the tension between the opposition in Hong Kong and the central government in Beijing is not completely unseen in other parts of the world where regional authorities have attempted to fight for more autonomy or even independence from the central governments (to name just a few cases, Quebec in Canada and Catalonia in Spain). Accordingly, studying the case of Hong Kong about whether a political camp's definition of the city's national identity may be correlated to its attitude towards asylum seekers can provide indications of how national identity may play a role in affecting the attitudes towards asylum seekers in general and how the politics of asylum seeking regime may develop in regions where the degree of the central government's authority is contested in particular.

Lastly, this thesis aims to test out if automated approaches to perform sentiment analysis on textual data in Chinese language can generate reliable results. van Atteveldt et al. (2021) note the lack of endeavours to utilise automated methods on non-English texts for sentiment analysis tasks which hampers the development of comparative research in the communication field, and yet automated sentiment analysis has the benefit of being easier to scale to larger datasets where human coding would become too time- and resource-consuming despite its high degree of validity. Although the two studies by Ng, Choi, et al. (2019) and Ng, Choi, et al. (2021) which used human coding to analyse the narrative on asylum seekers in Hong Kong by mass media between mid-2015 and mid-2016 can more thoroughly tease out the specific ways of how non-refoulement claimants are depicted besides knowing the sentiments of the articles, this approach may gradually become very laboursome if a study was to compare the reportage of asylum seekers between newspaper outlets over the span of several years. Accordingly, using the news articles from 2019 to analyse whether attitudes towards non-refoulement claimants may vary according to a newspaper's political affiliation via automated approaches can test the validity of these techniques, thereby checking whether machine-learning-based sentiment analysis is appropriate to be adopted on scaling the analysis to larger document corpora about how socio-political issues may be portrayed in Hong Kong as conditioned by certain features of the news articles or media outlets. If the results generated in this thesis via machine learning models are overall reliable, it will be worthwhile to test how such techniques may also fare in other types of

1. Introduction

policy documents such as speeches made by officials and legislators in the legislature in future research. In short, the validation of machine-learning-based methods in the context of social sciences research for Chinese text data can open up more possibilities for investigating intriguing questions.

1.2 Outline

This thesis will proceed as follow. In section 2, I will review the literature on the theories about how the perception of national identity might affect attitudes towards immigration, how each major political camp in Hong Kong perceives the national identity of Hong Kong, and how each camp perceives asylum seekers. Section 3 explains how the hypothesis is formulated and its expected direction, how the data was collected and pre-processed, what kind of models will be used for the analysis and how the variables will be operationalised. Section 4 records the process of the data analysis from preliminary exploring the data to modelling them, presents the results and discusses the findings. Section 5 concludes this thesis and mentions the potential implications of the findings. To briefly summarise the findings, it is predicted that newspaper outlets from the pro-Beijing camp, which emphasises on the ethnic Chineseness shared among local Hong Kong citizens and mainland Chinese for constructing the national identity of Hong Kong, are indeed more likely to portray asylum seekers more negatively than non-pro-Beijing newspapers after accounting for the presence of different topics and other features of a news articles.

Before moving on, it should be mentioned that I will use the words “asylum seekers”, “refugees” and “non-refoulement claimants” interchangeably. It is true that difference exists between the definitions of asylum seekers and refugees¹. Nevertheless, since the above three terms are generally used to refer to the same group of population in Hong Kong who flee from their own countries to escape from threats of torture, treating these terms as synonyms in this context should better cover how this group of population is portrayed by media outlets in different political camps. In fact, scholars researching on this issue often use the terms I listed without strictly differentiating them (e.g. Vecchio (2014), Ng, Choi, et al. (2019)).

¹For the complete documentation of the default parameters of the models used in this thesis, refer to the websites of scikit-learn and XGBoost Documentation — xgboost 1.5.1 documentation.

Chapter 2

Literature Review

To understand how the different conceptions of the Hong Kong identity may affect attitude towards asylum seekers, it is important to first review the literature on how national identity may affect attitudes towards immigrants and how Chineseness may be articulated on theoretical level, as well as how political camps in Hong Kong may define the city's national identity and the situation of asylum seekers on empirical level.

2.1 Theories

2.1.1 What is national identity?

I will adopt the definition proposed by Greenfeld and Eastwood (2007), which is how an individual relates to a broader societal order called the nation and is central to modern political and social organisations. Lewin-Epstein and Levanon (2005) point out modern nationalism's conception that people with similar cultures (e.g. language, myths, historic territory) should congregate together as the basis of political community called the nation-state. A closely related but not identical concept to nationalism is ethnicity, defined by Varshney (2007) as sense of belonging shared by a group in ancestry, language, history, culture and/or history. Whereas nations have their own territory-based political authority, this is not necessarily the case for ethnic groups.

Greenfeld and Eastwood (2007) also elaborate on some peculiar features of nationalism. For instance, the secular world where nations are located is deemed the most important to human beings and their life. But the more well-known characteristic is that the world is perceived to be divided into distinct communities (i.e. nations) occupying their own territories, and each member of "the people" within a nation is treated as equal to one another. It is also "the people" who

2. Literature Review

ultimately control political legitimacy and authority. These features together contribute to two fundamental implications. Socially, the fact that nationals are considered as being equal and that status is achieved instead of ascribed promotes social mobility. Politically, nationalism has made modern (nation-)states “impersonal” (p.261) because whoever holds the government office does not affect whatsoever the nation-state’s existence and legitimacy which ultimately rest in popular sovereignty (Greenfeld and Eastwood 2007). But for cultures that had been existing long before nationalism’s entry into these societies, the constructions of novel national identities may often rebrand pre-existing cultures despite the latter’s internal heterogeneity, of which Chineseness is an example which I will look into later.

Lastly, Greenfeld and Eastwood (2007) propose a typology of nationalism with two criteria. The first one is how the nation-state relates itself to its members, from the individualist end of treating a nation as merely a massive group of individuals to the collectivist end of which individuals are subservient to the nation. The other criterion is the well-known “ethnic” versus “civic” dimension which is extensively studied, including its relation with the perception towards immigrants which I will discuss later. In general, ethnic nationalism defines a nation’s members based on ascriptive and rather objective standards such as descent, language and customs which are largely decided at birth, whereas civic nationalism focuses less on the members’ cultural backgrounds than their voluntary allegiance to a political community and its legal institutions within a political territory (Greenfeld and Eastwood 2007; Lecours 2000; Lewin-Epstein and Levanon 2005; Heath and Tilley 2005).

A note on the ethnic-civic dimension is that these two aspects are not mutually exclusive since Medrano and Koenig (2005) highlight that nationalism in one country can often incorporate both civic and ethnic features, which is quite frequently observed from empirical. Accordingly, scholars often create other dimensions of describing nationalism theoretically. For example, Hjern (1998) proposes to additionally gauge whether identifications with ethnic and civic features of nationalism are high (“multiple national identity”) or low (“pluralist”) at the same time (p. 453). Meanwhile, some scholars prefer more fine-grained categories of citizenship criteria. Medrano (2005) breaks the ethnic and civic dimensions further into five categories while describing the importance of criteria for citizenship perceived by the Spanish population, namely (p.137),

- the civic dimension being broken down into:

2. Literature Review

- Pure republican criteria: Long-term residence, feeling of belonging, and being a citizen of that state
- Political criteria: Respect for laws and political institutions
- and the ethnic dimension into:
 - Territorial criteria: Place of birth
 - Cultural criteria: Language, religion, customs and traditions and
 - Descent

2.1.2 Impact of conceptions of national identity on attitudes towards immigrants

Perhaps one of the most prominent consequences of how national identity is perceived by individuals is their attitudes towards immigrants. This is because national identity simultaneously defines members and outsiders of a nation (Hjerm 1998; Lewin-Epstein and Levanon 2005; Heath and Tilley 2005). For simplicity's sake, I will mainly focus on how tendency towards defining national identity in ethnic and civic terms may be correlated with perceptions towards immigrants.

To begin with, although it is generally agreed that characteristics used to demarcate membership in a nation can be classified as ethnic- or civic-based on a macro-level, additional categories are often needed to capture the general patterns of how national identity is defined within a country and how such definitions may affect citizen's attitudes towards immigrants. A potential reason for such a need is that although individuals may hold a rather "pure" civic conception of identity mostly devoid of ethnic components, it is quite unlikely that individuals would consider ethnic components as standalone enough for constituting a nation's citizenship without any civic components. Indeed, empirical evidence from Hjerm (1998), Heath and Tilley (2005) and Medrano (2005) demonstrate that it is very rare for individuals to perceive national identity solely in ethnic terms, since those who attribute importance to ethnic-based components are also likely to do so to civic-based ones simultaneously. Often, scholars may even challenge the ethnic-civic-dimension in classifying the perception of national identity and propose alternative schema. Drawing on the Weberian idea of social closure which measures the number of obstacles imposed by a group for outsiders to join, Medrano (2005) proposes the "postnationalist" versus "credentialist" (p.136)

2. Literature Review

dimension which differentiates how restrictive people consider acquiring citizenship to their nation should be. Specifically, postnationalists impose few limits on foreigners to become citizens, whereas credentialists require a long list of criteria being fulfilled. That said, even though there exist multiple and equally valid alternatives for categorising how national identity may be defined, the ethnic-vs-civic dimension is quite broadly applicable to explain the typologies of nationalism in countries with some modifications and thus is a decent starting point.

As for the potential impact of national identity on attitudes towards immigrants, the literature tends to agree that the more an individual holds ethnic-based criteria of membership in a nation, the more likely (s)he will be averse against immigrants (i.e. xenophobic). It seems that the effect of ethnic-based national identity conception on xenophobic attitudes is consistent across different immigration policy regimes differing on the ease of naturalisation and the mode of integration. Hjerm (1998) discovers that despite Germany, Australia and Sweden having considerably different immigration policy regimes, individuals who agree that having both ethnic and civic characteristics are crucial for being a member of their nations are more likely to be xenophobic than others who only use civic components to demarcate nationhood or ascribe little importance to nationalism. Similar findings are observed by Heath and Tilley (2005) in the UK context, as British who place more emphasis on ethnic-based nationhood criteria will be more inclined to approve reducing immigrants inflows and more stringent measures against illegal immigrants. Interestingly, Heath also discovers that after controlling for how national identity is conceived, the extent to which an individual is attached to the nation is no longer significant in affecting his/her attitude towards immigrants.

In some cases, however, the effect of national identity on perceptions towards immigrants may be less due to whether individuals hold ethnic-based definitions of national identity than how many criteria individuals deem to be indispensable for obtaining a nationality. For instance, Medrano (2005) argues that even though there is an association between including ethnic-based (or in his terminology “ethno-biological republican”) (p.148) criteria for obtaining Spanish citizenship and holding negative attitudes towards immigrants, it is ultimately how many criteria a Spanish regards as important for being a Spanish citizen that better predicts attitudes towards immigrants, evidenced by the stronger correlation between these two items found in his study.

2. Literature Review

Furthermore, depending on the ethnic composition of a nation-state, not only may different ethnic groups have their own definitions of what constitute to the ethnic and civic components of national identity, but also the relations between perceptions of national identity and attitudes towards immigrants may go into rather unexpected directions. Lewin-Epstein and Levanon (2005) discover that in the Israeli society which can be broadly separated into three ethnic groups (namely, veterans Jews who have lived in Israel for a long time, Olim who are recent Jewish immigrants from former Soviet Union regions, and ethnic Arabs citizens in Israel), there are only moderate similarities in terms of which components should be included into the ethnic and civic dimensions of national identity, as they disagree on whether the ability to speak Hebrew and whether being born in Israel should be labelled as ethnic or civic components. Moreover, contrary to the expectation in the literature, veteran Jews who consider the Israeli nationality as more ethnic-based are actually *less* averse towards immigrants, possibly because they are likely to perceive immigrants as mostly overseas Jews who share the same ethnicity (Lewin-Epstein and Levanon 2005). Nevertheless, the surprising case of Israel is likely to be more due to its specific socio-political context rather the theoretical pitfalls of how ethnic-based conceptions of national identity may be related to xenophobia per se.

2.1.3 Chineseness in Hong Kong

Given over 90% of the population being ethnically Han Chinese by 2016 (Home_Affairs_Department 2018) and also large-scale pro-democracy political movements in the 2010s challenging the authority of the HKSAR and Chinese Governments (particularly the 2014 Umbrella Movement and the 2019 Anti-Extradition Law Protest), Chineseness indeed occupies a salient position on the discourse of the Hong Kong identity. It is then important to understand what Chineseness may mean in Hong Kong, a city under British colonial rule for over a century until 1997.

Although Chineseness is usually connoted with a homogeneous image of a group of people with the same ethnic origin, place of residence or cultural characteristics (e.g. language, customs etc.), Chun (1996) problematises this conception not least because expressions used to describe different aspects of Chineseness actually contains heterogeneity. For example, despite the contemporary use of Han Chinese to denote ethnic homogeneity, people under the Han empire were anything but ethnically homogeneous. It was not until the 1911 Revolution then China became a nation

2. Literature Review

inhabited by the single ethnic group called *Zhonghua Minzu*, and a homogeneous national culture was accordingly invented and legitimised by the state. The efforts to articulate a Chinese national identity continued after the foundation of the People's Republic of China (PRC) in 1949, as both the Kuomintang (KMT) regime in Taiwan and the PRC (after the Maoist period) extensively used Chinese history, political ideology and/or values to construct their national cultures and identities to claim themselves being the orthodox Chinese nation.

Whereas Chineseness was constructed by Chinese regimes holding political authority in Taiwan and mainland China, Chineseness manifested quite differently in Hong Kong where it was a Chinese society ruled by the British Government before 1997. According to Chun (1996), most inhabitants in Hong Kong had identified as Chinese until the establishment of the PRC in 1949. So (2015) also mentions that before 1949, ethnic Chinese in Hong Kong did not intend to permanently reside in the city and identified with the broader Chinese nation, evidenced by their participation in the anti-colonial Hong Kong-Canton general strikes in the 1920s. When the Chinese Communist Party (CCP) took power in 1949, however, the identity of ethnic Chinese in Hong Kong started to gradually depart from that on the mainland. As Hong Kong had become the venue to the KMT and the CCP's contest of the Chinese identity in the 1950s, the British Government then attempted to pull Hong Kong away from this nationalist conflict by actively promoting economic transformation from a Chinese entrepot to exported-oriented economy, and the colonial government started to create a new Hong Kong identity to further distance the citizens from Chinese nationalism following the outbreak of leftist riots in the late 1960s (Chun 1996; So 2015).

With intellectuals in Hong Kong paying more and more attention to local instead of Chinese affairs and the huge influx of illegal immigrants from China in the 1970s, the conflict between the old and newly arrived residents eventually led to the rise of a Hongkonger identity (So 2015). By mid-1980s, Hong Kong eventually developed its own culture which was indifferent to politics while also fusing habits from Chinese and Western cultures together (Chun 1996). As the transfer of the city's sovereignty approached in the 1980s, however, the Hong Kong identity became more complicated. On one hand, the investment opportunities in and relocation of factories to mainland China due to the reform and opening up policy helped resurge Chinese nationalism among Hong Kong business elites and working class who married mainland Chinese (So 2015). Meanwhile, the

2. Literature Review

uncertainty associated with the transition of sovereignty, especially after the Tiananmen Massacre in 1989, caused the public to be more conscious of values respected in Hong Kong such as the rule of law and civil liberties and thus increased their perceived difference between Hong Kong and China (So 2015; Fong 2017b; Lin and Jackson 2021) .

Amidst the fear of Hong Kong losing its unique after the handover to China in the 1990s, the *One Country, Two Systems* (OCTS) framework which stipulates Hong Kong's autonomy and continuity of the capitalist system for 50 years was accordingly put in place to secure confidence of the public regarding the handover (So 2015; Fong 2017b) . Indeed, the first decade after 1997, So (2015) and Fong (2017b) report that overall more and more Hongkongers identified as Chinese despite Beijing's increasing intervention (more on this later) after the mass protest against Article 23 in 2003, and such sentiment peaked in 2008 when China was the host of that year's Olympic Games. But the Hong Kong identity has once again become prominent since 2009 as Hongkongers become aware of Beijing's increasing political interference and the social disruption brought by the large influx of mainland Chinese, and Fong (2017b) highlights how the Hong Kong identity has emerged in a mentality of resisting against "invaders" from the mainland during the period (p.19).

In short, Chineseness as how closely Hong Kong people identify themselves as members of the Chinese nation have evolved throughout the course of history, and one could argue that following the transferral of Hong Kong's sovereignty to China, the Chineseness of Hong Kong identity has become more and more contested. This phenomenon is best reflected by how each prominent political camp in Hong Kong envisions the Hong Kong identity.

2.2 Empirics

2.2.1 National identity of Hong Kong defined by different political camps

With Hong Kong's major political cleavage being demarcated by the city's relations with Beijing and the pace of democratisation, different political camps may adopt their own vision of what Hong Kong's national identity should be. Indeed, Lecours (2000) highlights the importance of political goals in affecting how political actors with different ideologies on a region's relations with the national government may formulate national identity through the case study of Quebec. He argues that the secessionist Parti Quebecois (PQ) is motivated to adopt an ethnic-based

2. Literature Review

nationalism which emphasises on French being the Quebecois' mother tongue because the party needs a more salient cultural marker to convince Quebecois that they are indeed qualitatively different from the Anglophone Canada and thus warrants a separate nation. By contrast, since the autonomist Parti Liberal du Quebec's (PLQ) goal is to demand more political autonomy instead of secession, the party can adopt more civic-based framing of nationalism which does not restrict membership to the Quebecois society exclusively to Francophones while not explicitly relying on ethnic markers to distinguish between Quebecois and Canadians.

In the following paragraphs, I will discuss the political goals of the three prominent political camps in Hong Kong regarding the city's political regime and relations with the Beijing Government, namely, the pro-Beijing camp, the pro-Democracy camp and the newly arising localist camp in the 2010s, and how their political goals may affect their articulation of Hong Kong's national identity.

The pro-Beijing camp

As the politically conservative force governing Hong Kong, the pro-Beijing camp's objective can be summarised as legitimising China's sovereignty over the city and protecting the political interests of China. Members of the camp include the HKSAR Government itself, political parties and societal actors who lean towards the Beijing Government etc. Perhaps the most prominent manoeuvre by the pro-Beijing camp to achieve the above objective is counteracting against demands for more rapid democratisation in both the executive and legislative branches to prevent the opposition from gaining control over the political institutions (Ma 2012) .

Apart from controlling high-level political institutions in the HKSAR, the pro-Beijing also invests in efforts to foster patriotism towards China in different societal sectors, of which education is one. Bottom up endeavours to instil the Chinese identity into students' minds are organising trips to China as well as pedagogy of pro-China ideologies and Chinese cultures, by pro-Beijing schools and regular schools receiving sponsorship from the Government or pro-Beijing actors, whereas top-down endeavours from the Government mainly revolves around policies on national education such as teaching Putonghua in primary school since 1998, encouraging schools to raise the Chinese flag and sing the Chinese anthem in significant occasions and the failed attempt to make national education itself a mandatory subject which avoids discussion on sensitive issues

2. Literature Review

in 2012 (Lau 2013; Fong 2017b) . The pro-Beijing camp also sets up support organisations in communities for electoral mobilisation, as Loh (2010) and Fong (2017b) out how the largest pro-Beijing parties in Hong Kong have numerous subsidiaries grassroots organisations which can serve to mobilise for votes during elections under the coordination of the Liaison Office (LOCPG), and Fong (2017a) reports an alleged vote-rigging incident in 2015 during which pro-Beijing-affiliated elderly homes were spotted sending elderlies, often of low cognitive abilities, to cast their votes.

With the pro-Beijing camp's goal of securing the ultimate authority and sovereignty of the Beijing Government over the HKSAR, its vision on how the identity of Hong Kong should be is then closely intertwined with the Chinese identity. In other words, pro-Beijing forces view the Chinese and Hong Kong identities as a whole in which Hong Kong is subordinate to the Chinese nation, and identification with the Chinese nation is translated into support towards the incumbent PRC regime, which self-proclaims as the “vanguard of the Chinese people and the Chinese nation” (*Full Text of Constitution of Communist Party of China - Xinhua* 2017) . Indeed, Lin and Jackson (2021) argue that at least during 2012-19, the HKSAR government primarily portrays the Chineseness of Hong Kong citizens in an essentialist and narrow ethnic-based manner which emphasises on the shared biological inheritance from the ancient Peking Man between Hong Kong citizens and mainland Chinese while downplaying other aspects such as values, cultural heritage and lifestyle. For instance, former Chief Executive CY Leung and pro-Beijing politician Ronny Tong claimed that Hong Kong citizens were obliged to identify themselves as Chinese due to their ascribed biological similarities with mainland Chinese. Ethnic markers were also co-mentioned while HKSAR officials were describing the PRC, as former Chief Secretary Matthew Cheung stated that he was proud as “a son of the Chinese” while witnessing the progress made by mainland China (Lin and Jackson 2021) (p.914).

This view echoes with that held by the Beijing Government, as Chinese President Xi Jinping once claimed that Hong Kong's fate was closely linked to mainland China, and the city was crucial for achieving the Chinese dream (Lin and Jackson 2021) . In essence, as the ruling government of post-1997 Hong Kong is now under the sovereignty of the PRC, it is likely that the HKSAR Government and its political ally, the pro-Beijing camp, actively promote Chineseness as an ethnic feature that both Hong Kong citizens and mainland Chinese share closely to realign the Hong Kong identity, which is deemed to had developed separately in British Hong Kong (Chun 1996;

2. Literature Review

So (2015) , under the umbrella of Chinese so that Beijing could exert more control over the city's ideology. By contrast, civic values are less salient in the official discourse in promoting Chineseness in Hong Kong (except for the value of patriotism which is ultimately another manifestation of the ethnic conception of Chineseness), but are often used by the pro-Democracy camp to portray Hong Kong's distinct (but **not** independent) identity from mainland China.

The pro-Democracy camp (or pan-democrats)

According to Ma (2012) , even though several pro-democracy political groups were founded in the 1980s after the commencement of the city's democratisation and elections as well as the signing of the Sino-British Joint Declaration in 1984, it was not until the late 1980s when these groups congregated together for better mobilisation power in large-scale political movements, especially during the 1989 Tiananmen Movement when the Alliance for the Support of the Patriotic and Democratic Movement (ASPDMC) was formed to support students protesting for democracy in Beijing. Shortly afterwards, the first pro-democracy party, *the United Democrats of Hong Kong* (the predecessor of the present Democratic Party) was created in 1990 for next year's first direct election in the legislature (Ma 2012) . After the handover in 1997, the pro-democracy camp remains as one of the significant political forces in the HKSAR, including the provision of an alternative perception of the Hong Kong identity vis-a-vis the official rhetoric.

Compared to the pro-Beijing camp which considers protecting the interests of the Chinese nation (currently under the PRC regime) as paramount over other concerns, the pro-democracy camp instead prioritise values such as the rule of law and civil liberties which are regarded to be constitutive to the civic elements of Hong Kong's identity (Kwan 2016) . This is evidenced by the pro-democracy camp deeming the Article 23 legislation back in 2003 as contravening personal freedom under the pretext of national security and thus openly opposed to and organised the 1 July rally against the law's promulgation (Ma 2012) , and the camp also supported the Occupy Central Movement (a precursor to the 2014 Umbrella Movement) to protest against Beijing's proposal of only allowing pre-approved candidates to run for Chief Executive elections in August 2014 as violating democratic principles of universal suffrage (Kwan 2016) . It can thus be argued that the pro-Democracy camp puts more importance on protecting the civic values that they

2. Literature Review

deem crucial for the identity of Hong Kong, even if this means often clashing with the interests of Beijing.

When it comes to identifying which nation Hong Kong belongs to, however, it should be noted that the pro-democracy camp in general still identifies themselves as Chinese, as Kwong (2016) states that especially for the older generation pro-democratic politicians, they are in favour of Hong Kong's sovereignty being transferred from Britain to China in 1997, and they claim to share affinity with China's history, culture and ethnicity despite opposing the CCP's one-party dictatorship. In fact, the identification with the Chinese nation by the pro-democracy camp is well-inscribed in the manifesto of the largest and most historic pro-democracy party, *the Democratic Party*, as the very first article clearly claims that "Hong Kong is an indivisible part of China" (Manifesto 2021). The pro-democracy camp's identification as members of the Chinese nation may very well explain how they contextualise Hong Kong's democratisation movement and why they do not think Hong Kong identity is independent of the Chinese one. Kwong (2016) mentions that the pro-Democracy camp generally considers Hong Kong's democratisation as integral to that of mainland China, and the annual candlelight vigil by the ASPDMC is a totem of this mentality since some organisers perceive a democratic China, if realised, might subsequently bring democracy to the city as well.

Admittedly, the pro-democracy camp attempted to appeal to localist supporters when the ideology gained huge popularity around the time of the 2016 Legislative Council election, such as calling for the right of self-determination for Hong Kong (Kaeding 2017; Fong 2017b). Nevertheless, this does not mean the pro-democracy camp giving up on caring about political issues in China while relating them to the situation in Hong Kong. Even in the 2021 Tiananmen Massacre candlelight vigil, the ASPDMC stated that Hong Kong people shared "a common fate with those oppressed and deprived of freedom in China", and the organisation would struggle until both Hong Kong and China became democratic (ASPDMC 2021).

In short, the pro-democracy camp considers civic values in Hong Kong such as the rule of law and civil liberties to be constituting to the distinctiveness of Hong Kong and is averse towards Beijing's political interventions which often run against these values. Nevertheless, the camp still regards Hong Kong as belonging to the same nation with mainland China and is even willing to help their Chinese compatriots in political movements which advocate freedom and democracy.

2. Literature Review

The geographical scope of how wide Hong Kong's political movement to democratisation should concern, then, is partly what makes the localist camp different from the pro-democracy camp even though both favour a democratic regime in Hong Kong.

The localist camp

Broadly speaking, the rise of localism as a third prominent political ideology in the 2010s was due to both the increasing integration with mainland China both politically and socio-economically as well as the perceived incompetence of the pro-Democracy camp to push forward substantive democratisation in Hong Kong. Although the term "localist camp" in reality represents an aggregation of groups which vary considerably in choosing violent or non-violent means of resisting against the Beijing Government, whether Hong Kong should abandon OCTS and pursue independence, and are much less coordinated in their actions compared to the pro-Beijing and pro-democracy camps, organisations belonging to this camp hold strong aversive sentiment against China **both** in terms of the PRC regime and mainland Chinese who are deemed to be depriving the locals of already scarce public resources or disrupting the way of life of ordinary Hong Kong citizens, and they (Kwong 2016) . The first substantive discourse to materialise localism as a political ideology was "Hong Kong as a city-state" published in 2011 by scholar *Chin Wan-kan* which, among many other ideas, advocates separating the integration between Hong Kong and mainland China as well as prioritising the needs of Hongkongers (So 2015) . The following paragraphs will outline how the rise of localism can be treated as a backlash against both long-existing political camps in Hong Kong while also exploring how localists perceive Hong Kong's national identity.

Starting with the so-called "China factor", the localist camp obviously rejects the pro-Beijing camp's (and the Chinese Government) claim that Hong Kong's identity should be treated as one with that of Chinese. Fong (2017b) illustrates how the China factor has facilitated the rise of localism under the framework of "state-building" versus "peripheral" nationalism (p.2), where Hong Kong is attempting to consolidate its unique identity against Beijing's incorporation of the city into the hegemonic national culture. Specifically, after 500,000 Hongkongers had rallied against the Article 23 legislation in 2003, Beijing then started to increase its interventions in

2. Literature Review

multiple areas, such as asserting that its authority overrides the city's autonomy, sponsoring pro-Beijing candidates during elections and fostering further economic integration between Hong Kong and China in finance, tourism and trade.(So 2015; Fong 2017b) . In turn, the high proportion of respondents in a 2016 survey identifying themselves as Hongkongese amidst the foundings of more radical localist organisations which aspire to make Hong Kong a separate nation can be evidence of how the PRC's attempts to assimilate Hong Kong into the hegemonic Chinese identity has risen the consciousness of Hong Kong people about their uniqueness vis-a-vis China and thus foster the increasing popularity of localism (Fong 2017b) . Apart from Chin's foundational discourse on localism as mentioned before, overt advocacy of Hong Kong becoming an independent nation from China is often proposed by some localist organisations, most signified by the establishment of the disbanded *Hong Kong National Party* in 2016 by Andy Chan which aimed to build a "Hong Kong Republic" (Kaeding 2017) (p.165). Some organisations even dare openly insult the PRC to challenge its authority and legitimacy, as two former legislators from the self-determination-advocating *Youngspiration* had taken their office oaths by pronouncing "China" derogatorily in 2016, which caused them to be disqualified from their seats shortly afterwards (Kaeding 2017) .

Due to the localist camp's dissociation of Hong Kong with mainland China concerning both Beijing's assimilation and which nation Hong Kong belongs to, they also reject the pro-Democracy camp's idea that political development of Hong Kong should be linked to that of China. In effect, the indifference of the localist camp to China's political situation is largely due to the "transition fatigue" which witnesses increasing dissatisfaction with the traditional pro-democracy camp has started to accumulate due to the lack of substantive progress in democratisation by 2008 (Kwong 2016) (p.63). The Democratic Party's decision to negotiate with the Beijing Government for the political reform package and its subsequent support in May 2010 created profound division within the pro-democracy camp as certain parties believed that more radical measures were needed to fend off Beijing's interference into the city, but the more significant turning point for localism to gain its popularity was the 2014 Umbrella Movement when some youngsters blamed the pro-democracy camp for the Movement's failure (Kwong 2016; Kwan 2016) .

The disagreement between the pro-democracy and localist camp lies less in what kinds of values Hong Kong should uphold (as both camps aim at fighting for democracy and autonomy from Beijing) than whether Hong Kong should involve in China's democratisation. Specifically,

2. Literature Review

the pro-democracy camp's insistence on Hong Kong's responsibility to help build democracy in China is met with strong criticism from the localists, who consider that Hong Kong should focus only on its political development since binding Hong Kong's future to a democratic China would hamper the city's own interests (Kwong 2016; Kwan 2016). Thus, the different conceptions of which nation Hong Kong belongs to between the two camps and how much Hong Kong should dissociate from China have caused both camps to have rather contentious relations, as some pan-democrats also consider the localist camp's hostility towards mainland Chinese as xenophobic (Kwong 2016).

In short, the localist camp does not just treat the Hong Kong identity as distinctive but ultimately belongs to the Chinese nation like the pro-democracy camp. Instead, Hong Kong is a separate political entity that should avoid close integration with mainland China in order to protect its own interests. This causes the localist camp to become a third force in Hong Kong's political arena in the 2010s confronting against both the pro-Beijing and pro-Democracy camps.

Before moving onto hypothesising how the three camp's different views on Hong Kong's national identity may impact their views on asylum seekers within the city, table 2.1 summarises how they differ on the vision of the Hong-Kong-China relations.

Table 2.1: Comparison of the three camps on key dimensions of Hong-Kong-China relations and their attitudes towards non-refoulement claimants (the last row)

	Pro-Beijing camp China	Pro-Democracy camp China	Localist camp Hong Kong
The nation that Hong Kong belongs to			
Attitude towards political development in China	Political ally of the ruling PRC	Opposes one-party rule of the CCP, sympathetic to Chinese people	Indifference, only focuses on Hong Kong's situation
Prioritised political values	Patriotism and nationalism over individual rights	civil liberties, the rule of law and political autonomy	those of the pro-democracy camp plus prioritising Hong Kong's interests
Means of contention	N/A (as an ally of the Government)	Non-violent	Both violent and non-violent
Degree of within-camp coordination	Strong	Moderate	Weak
Attitude towards non-refoulement claimants in Hong Kong	Hostile	Sympathetic	Undetermined (due to the lack of information)

2.2.2 Hong Kong's non-refoulement policy regime

Regarding the legal framework and policies of assessing asylum claims and refugee recognition, the Hong Kong Government has a strong position of not making the city a destination for refugee resettlement. Although Hong Kong is a signatory of the 1992 United Nations Conventions Against Torture (UNCAT) which requires the Government to temporarily house asylum seekers and process torture-related asylum applications, the city has never signed the 1951 Refugee Convention and thus has no legal obligations to permanently resettle asylum seekers and refugees within its territory (Ng 2020) . This remains true even after the implementation of the Unified Screening Mechanism (USM) in 2014 which made the Government become the sole handler of all the non-refoulement claims, including those which had been previously handled by the UNHCR for refugee status (Ng 2020; *Unified Screening Mechanism Procedures* 2021) . Instead, the Government reiterates its lack of intention of determining individuals' refugee status and allowing refugees to permanently settle in the city, while also claiming that the recognition and resettlement of refugees should still be referred to the UNHCR "in accordance with its mandate" (Department 2021b) , even though the UN agency states on its website that it would no longer process new refugee and asylum claims following the implementation of the USM (*Unified Screening Mechanism Procedures* 2021) . Simply put, the official institutions responsible for assessing asylum applications in Hong Kong hardly have intention of making the city a destination of permanent resettlement for individuals fleeing from their countries because of persecutions.

Even if the HKSAR Government is obliged to adjudicate non-refoulement applications as per the UNCAT, the legal and social status of the asylum seekers are quite precarious while awaiting their application results. To begin with, individuals must wait until their visas expire in order to be eligible for non-refoulement applications and then have to report regularly to the Immigration Department (*Unified Screening Mechanism Procedures* 2021) , an arrangement that Ng (2020) considers as the Government's attempt to exclude asylum seekers from the rights granted to legal residents in the city (e.g. working,) and receiving welfare supports through criminalisation. Furthermore, the Immigration Department has the final authority to prosecute the asylum seekers for having illegally overstayed beyond the validity of their visas, and officials frequently call these populations as *illegal immigrants* (often with racial labels) (Ng 2020) . Coupled with the Government's refusal to recognise and resettle refugees in its territory, the institutional framework

2. Literature Review

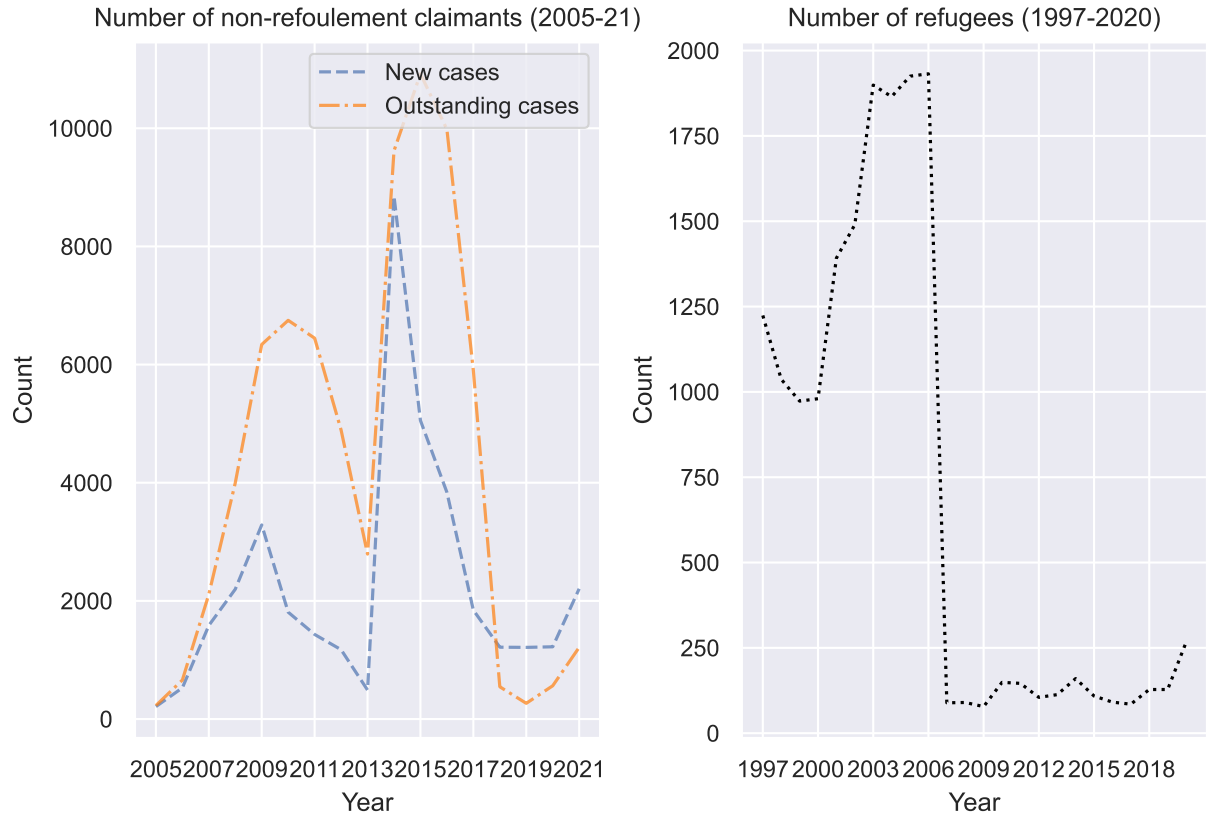


Figure 2.1: Number of non-refoulement claim cases (left) and refugees (right) in Hong Kong

of asylum seeking in Hong Kong can thus be described as unwelcoming towards those who are fleeing from persecutions in their countries of origin, since these applicants not only have to first become illegal immigrants for eligibility to request non-refoulement, but also they are under the constant threat of being charged for their immigration status.

The HKSAR Government's unwelcoming stance towards asylum seekers and refugees is likely to be one of the major reasons for the low numbers of non-refoulement claims and refugees in the city since 2005. For instance, the left plot in figure 2.1 (Department 2021a) shows that the number of *new* non-refoulement claims submitted each year from 2005 to 2021 on average was lower than 4000, and the sudden rise of new applications to around 9000 instances in 2014 was more due to the implementation of the USM requiring the Government to also assess claims previously handled by the UNHCR than the commencement of a more liberal asylum-seeking regime (Department 2021a). Likewise, the number of *outstanding* non-refoulement cases by year during the same period generally surpassed that of new cases except from 2018, a few years after

2. Literature Review

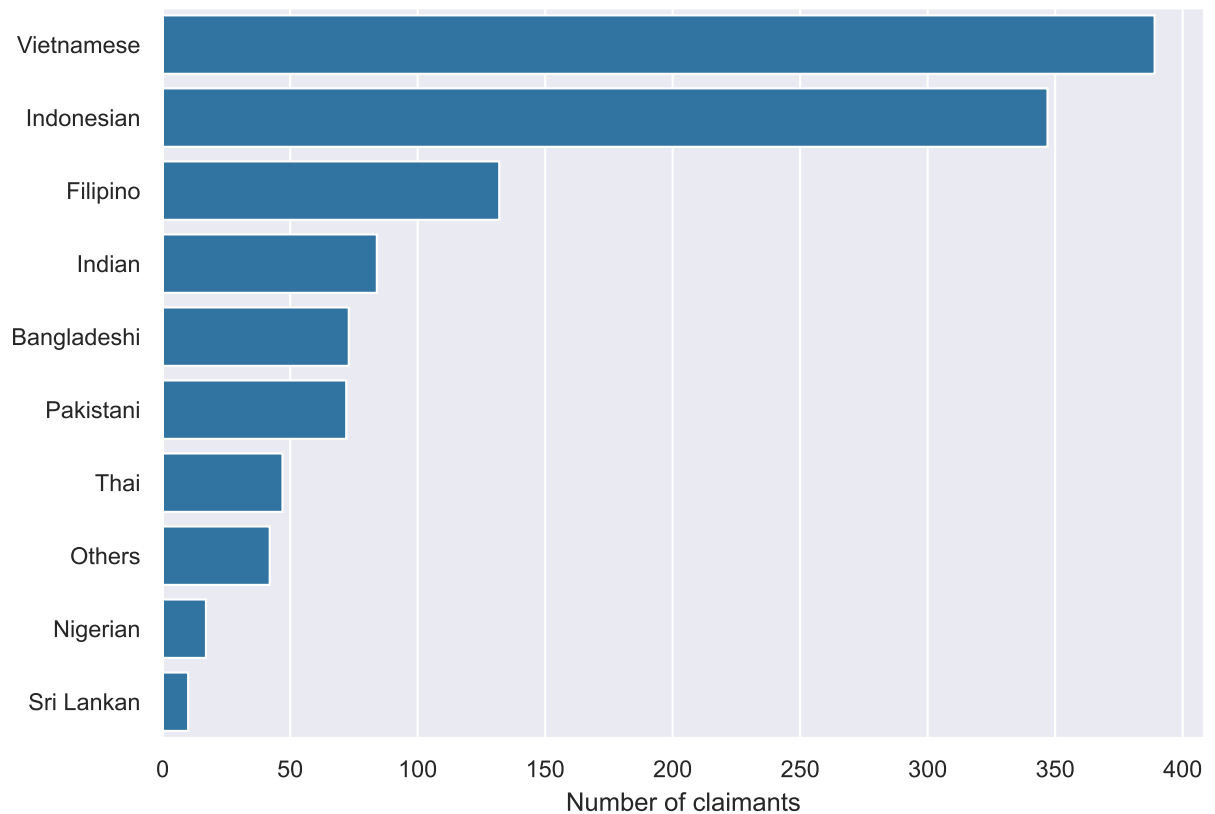


Figure 2.2: Number of non-refoulement claimants by nationality in 2021

the USM had been implemented. By 2018, non-refoulement claimants constituted to a mere 0.2% of Hong Kong's total population of over 7 million, a figure considerably smaller when compared to a total of 3.5 million refugees in Asia as Ng (2020) contrasts. The right plot in figure 2.1 (*Refugee Population by Country or Territory of Asylum - Hong Kong SAR, China / Data 2021*) shows how the number of *refugees* in Hong Kong has dramatically reduced to no more than 250 persons each year (save for 2020) since 2007 after the Government had closed the last refugee camp hosting Vietnamese who fled from the Vietnam War in the 1970s (Ng 2020) . Lastly, figure 2.2 (Department 2021a) indicates the nationalities of non-refoulement claimants, who mostly are from South and Southeast Asia, with a few from Africa as well. One particularly interesting insight from this graph is that despite the frequent racial labelling of non-refoulement claimants as South Asians by some media outlets (Ng, Choi, et al. 2019) , it is actually **Southeast Asians** who constitute the largest proportion of non-refoulement claimants.

2.2.3 How do different political camps respond to the issue of asylum seekers in Hong Kong?

After laying out the definitions of Hong Kong's national identity by various political camps as well as the policy regime of non-refoulement claims, how can we expect the attitudes towards asylum seekers may vary by political camp in Hong Kong? Do preliminary evidence align with expectations based on the literature on the relationship between the perception of national identity and attitudes towards immigrants?

Theoretical expectations in the case of Hong Kong

Starting with the pro-Beijing camp, because of its emphasis on the Hong Kong identity being a subset of the Chinese nation largely due to the ethnic homogeneity between residents in Hong Kong and on mainland China, I expect that this camp will be the least likely to express a positive attitude towards non-refoulement claimants who, as figure 2.2 shows, are considerably non-ethnic Chinese. This is because as research by Hjerm (1998) and Heath and Tilley (2005) demonstrate, more ethnic-based definitions of national identity are correlated with high degrees of aversion against immigrants, and in the case of the pro-Beijing camp, the significance of the ethnic Chineseness in defining the Hong Kong identity is so primordial that such rhetoric is propagated from pro-Beijing politicians in Hong Kong up to *Xi Jinping* himself as the President of the sovereign state of Hong Kong (Lin and Jackson 2021). One potential mechanisms for the pro-Beijing camp and its affiliated media outlets to be more hostile against non-refoulement claimants may be the conception of the majority ethnic-Chinese population in society as a family mentioned by Erni (2012), which potentially marks the non-refoulement claimants for permanently being strangers of the society due to their ethnic distinctiveness.

By contrast, even though the pro-democracy camp identifies with the Chinese nation, the fact that they also put considerable emphasis on civic values such as the rule of law and civil liberties for the definition of the Hong Kong identity is likely to make this camp at least less hostile towards asylum seekers in Hong Kong. In fact, given this camp's past history of activism related to human rights (most notably the annual candlelight vigil on 4 June to commemorate the 1989 Tiananmen Massacre), they may even be more sympathetic towards non-refoulement claimants who are, by

2. Literature Review

definition, forced to seek refuge overseas to escape from human rights violations (i.e. tortures) and demand that the HKSAR Government treat this group of population humanely.

As for the localist camp, although they argue that the interests of native Hong Kong citizens should be prioritised, they are usually referring to how Hong Kong citizens should have access to public resources before mainland Chinese who have just immigrated to Hong Kong recently. When it comes to ethnic minorities in Hong Kong, the localist camp has not mentioned whether the interests of native ethnic Chinese population should be prioritised. Therefore, it will be difficult to expect a priori with which attitude the localist camp and its associated media outlets may tend to view the issue of asylum seekers.

Empirical observations

Putting aside the lack of a clear expectation of the localist camp's attitude towards non-refoulement claimants, it seems that drawing from preliminary evidence, how the pro-Beijing and pro-democracy camps tend to view asylum seekers in Hong Kong are consistent with the expectations formulated according to the literature. In other words, the pro-Beijing camp in general creates an antagonising narrative against non-refoulement claimants, whereas the pro-democracy camp is more inclined to advocate for better treatment towards this group of vulnerable population.

Starting with the pro-Beijing camp, some politicians echo with the HKSAR Government's restrictive policies of non-refoulement claims by openly questioning the legitimacy of individuals seeking refuge from tortures in Hong Kong. For instance, pro-Beijing legislator Holden Chow submitted a motion in November 2016 under the banner of "combating bogus refugees" who worked illegally and committed crimes while awaiting for the decisions of their applications and called for the revision of the USM as well as the implementation of measures against illegal immigrants adopted by other countries (Yuen 2016). Some pro-Beijing parties also reach out to the public to sway the latter's opinion, such as the Democratic Alliance for the Betterment and Progress of Hong Kong (DAB) setting up a booth in the ethnic-minority-concentrated Tsim Sha Tsui district to hand out pamphlets about measures against "fake refugees" (Cheung and Grundy 2016). Perhaps the most dramatic proposal of dealing with the so-called "fake refugee" issue by the pro-Beijing camp was the former Chief Executive *CY Leung's* suggestion to withdraw from the UNCAT in 2016 so that Hong Kong would not be obliged to assess non-refoulement

2. Literature Review

claims anymore (Ng, Choi, et al. 2019) . Coupled with the high frequency of pro-Beijing media outlets in pejoratively portraying asylum seekers in Hong Kong as mentioned earlier, the pro-Beijing camp promotes an unwelcoming rhetoric against asylum seekers that is consistent with the Government's refusal to permanently resettle refugees and asylum seekers.

As the pro-Beijing camp advocates more stringent measures against non-refoulement claimants whom they deem as detrimental to the security of Hong Kong, the pro-democracy camp is considerably more sympathetic towards asylum seekers. When Holden Chow's motion was tabled for voting in the legislature in December 2016, it was barely rejected due to the pro-democracy camp's opposition of Chow's motion as "hypocrisy and discrimination in disguise" by indiscriminately applying the proposed "draconian measures" on all asylum seekers (Cheng 2016a) . Among all pro-democracy legislators, Dr Fernando Cheung was perhaps the most vocal in fighting for better treatments of asylum seekers in Hong Kong, as he met with non-refoulement claimants directly on several occasions while receiving the latter's approval (Justice Centre Hong Kong 2014; Vision First 2015; Refugee Union 2018). Indeed, legislator Cheung's prominent advocacy for the rights of asylum seekers indirectly caused him to become a victim of political smearing a few months before the 2016 Legislative Council election, as political banners impersonating legislator Cheung's party were seen hanging in streets which maliciously called the legislator "the father of refugees" (Ngo 2016). It should be noted that, nevertheless, a more lenient attitude towards asylum seekers does not mean that the pro-democracy camp totally disregards the perceived issue of bogus refugees. Even though legislator Charles Mok condemned Holden Chow's motion against "bogus" refugees as "disturbing", he also blamed the HKSAR Government for the existence of "bogus non-refoulement claims" due to the latter's lack of efficiency and lack of knowledge in screening out bogus claims which in turn exacerbated the issue (Cheng 2016a). But in general, the pro-democracy camp supports a more humane treatment towards non-refoulement claimants who are genuinely in need of fleeing from their countries due to threats of tortures.

Lastly, anecdotal evidence suggests that the localist camp at least adopts a less unwelcoming stance towards non-refoulement claimants compared to that adopted towards mainland Chinese. As one of the few localist lawmakers who were not disqualified in the oath-taking controversy by December 2016, Dr Cheng Chung-tai's speech which opposed Holden Chow's anti-fake-refugees

2. Literature Review

motion had two points which are worth highlighting¹ (*Hui4yi4 Guo4cheng2 Zheng4shi4 Ji4lu4* 2016) (p.1232-33). Firstly, he shared the pro-democracy camp’s view that the major reason for the problem of “bogus non-refoulement claims” is that the Government has not allocated sufficient resources to efficiently assess the claimants’ applications after the commencement of the USM in 2014. But the more intriguing point of view is how he framed Hong Kong’s signatory status of the UNCAT as a proof of the city having its own *de-facto* sovereignty from mainland China, and thus he opposed CY Leung’s suggestion of withdrawing from the treaty since it would signify Hong Kong’s backsliding in civilisation. Interestingly, the localist camp’s reservation of adopting more drastic measures against asylum seekers (e.g. quitting the UNCAT) has drawn smear from some pro-Beijing medias, such as an editorial from the Sun (2016) accused the localist camp of “remaining silent” on non-refoulement claimants wasting the public resources of Hong Kong. That being said, it is rather difficult to conclude how the localist camp views the issue of non-refoulement claimants from this single anecdotal incident.

¹For the complete documentation of the default parameters of the models used in this thesis, refer to the websites of scikit-learn and XGBoost Documentation — xgboost 1.5.1 documentation.

Chapter 3

Methods

In the previous chapter, I covered the literature on how perceptions of national identity may affect the attitudes towards immigrants, how political camps in Hong Kong define national identity and their attitude towards the treatment of non-refoulement claimants. With the above theoretical basis of how each political camp is likely to perceive the issue of asylum seekers in Hong Kong, I will now move on to investigate how informative the political affiliation of media outlets is for predicting the sentiments of news articles on non-refoulement claimants after accounting for other factors, such as the content and the metadata of the news articles. This chapter will discuss the research strategies in detail before presenting the analysis results in the next chapter.

3.1 Hypothesis

Following the aforementioned discussion, the pro-Beijing camp expresses more overt hostility towards non-refoulement claimants in Hong Kong than other political camps by frequently referring to the latter as “fake refugees” which warrant solutions derived by the HKSAR Government in order to protect the interests of the Hong Kong society. Therefore, I decided to formulate H_1 as follows:

H_1 : Pro-Beijing media outlets are more likely to negatively portray asylum seekers than media outlets from other political camps.

Should the analysis result demonstrates that the probability of a news article being negative (how the sentiments are operationalised will be discussed later) is higher if it is published by pro-Beijing media outlets, then this hypothesis should be accepted.

3.2 Data collection

The data will be drawn from Chinese language news media reports on asylum seekers in Hong Kong published by local printed newspaper outlets during 2019¹. For the retrieval of the relevant newspaper articles, I will rely on the *WiseNews* database which can fetch news articles with search keywords². While it would have been more ideal to also include online news media into the dataset as Lee (2018) mentions that they have become quite prominent in the 2010s for providing counter-narratives to the mainstream media which are often co-opted by the Chinese Government³, WiseNews does not directly provide the texts of online media news article and thus requires additional steps of web scrapping which are, unfortunately, beyond the author's capacity by the time of writing. That being said, including only the printed news media during 2019 would still cover 16 of the well-known news media which were included in the 2019 media credibility survey by CUHK (2020). In other words, just searching for printed news media outlets still allows me to discover whether the attitudes of reporting on asylum seekers by some of the most well-known and widely circulated media outlets in Hong Kong might be related to their political associations.

The reason for choosing to collect media reports published in 2019 is because it was the year when the Anti-Extradition Law Amendment protest first broke out in June and then lasted for several months, a period when one can argue the manifestation of political ideologies and identities should be the most salient for all the political camps. This should indeed be the case for both the pro-democracy camp whose supporters participated in several large-scale protests and sometimes even engaged in numerous violent clashes with the Hong Kong police to oppose a law that would have permitted extradition of fugitives to mainland China, and the pro-Beijing camp whose authority was once again immensely challenged after the 2014 Umbrella Movement. Accordingly, it would be interesting to investigate whether the issue of asylum seekers still remained its salience

¹For the complete documentation of the default parameters of the models used in this thesis, refer to the websites of scikit-learn and XGBoost Documentation — xgboost 1.5.1 documentation.

²The code for tuning the model can be found in the appendix.

³Hyper-parameters of the tuned XGBoost model used here are: {'objective': 'multi:softprob', 'use_label_encoder': False, 'base_score': 0.5, 'booster': 'gbtree', 'colsample_bylevel': 1, 'colsample_bynode': 1, 'colsample_bytree': 0.6000000000000001, 'enable_categorical': False, 'gamma': 0.5, 'gpu_id': -1, 'importance_type': None, 'interaction_constraints': '', 'learning_rate': 0.37, 'max_delta_step': 0, 'max_depth': 5, 'min_child_weight': 3.0, 'missing': nan, 'monotone_constraints': '()', 'n_estimators': 40, 'n_jobs': 16, 'num_parallel_tree': 1, 'predictor': 'auto', 'random_state': 1, 'reg_alpha': 0, 'reg_lambda': 12.0, 'scale_pos_weight': None, 'subsample': 0.7000000000000001, 'tree_method': 'exact', 'validate_parameters': 1, 'verbosity': None, 'eval_metric': 'mlogloss'}. To reproduce this model, simply create an XGBClassifier instance, copy the above hyperparameters as a dictionary and then use the .set_params(**dict) method on the XGBClassifier instance.

3. *Methods*

and was reported in similar manners given the co-occurrence of another arguably more salient political event (i.e. the anti-Extradition Law protests), as compared to when the issue first had become prominent back in 2016. Specifically, analysing newspaper articles in 2019 provides an opportunity to observe whether and how asylum seekers in Hong Kong might be represented by the media in tandem with major political events.

It is worth discussing some limitations of using media reports as the source of data. For starters, media reports do not necessarily reflect how individuals identifying with each political camp might perceive asylum seekers directly and might not be fit for investigating the causal mechanism of how definition of Hong Kong's national identity may cause a political camp to frame asylum seekers in particular ways in face of the public. Despite these shortcomings, media reports can still adequately help test the hypothesis of whether narratives on asylum seekers in Hong Kong may vary by political camps which share drastically different definitions of Hong Kong's national identity because media reports will likely reflect each political camp's preferred version of framing on the issue to which they would like to propagate. In fact, previous research about the framing of asylum seekers by mass media (Cooper et al. 2017; McKay et al. 2011; Ng, Choi, et al. 2019; O'Doherty and Lecouteur 2007) point to its potential impact on affecting how the public may perceive the issue via being a major provider of related information, and sometimes media reports may even serve to justify policy-making decisions or reflect the government's position on the issue (when the media simply repeat the information provided by officials). Drawing from Fong (2017a) and Lee (2018), the media industry in Hong Kong is quite remarkable for their political stances, and numerous outlets have their owners being affiliated with the Chinese and/or HKSAR Governments. Therefore, reports published by Hong Kong media outlets should be able to reflect political camps' attitudes towards asylum seekers and, perhaps more importantly, the version of framing that they would like to persuade their readers to accept, thereby making them a suitable data source to gauge how may each political camp's discourse on asylum seekers vary given their contrasting definitions of the city's national identity.

Language-wise, including only Chinese-language news media omits how asylum seekers may be represented by English-language media in Hong Kong (e.g. the South China Morning Post) and thus cannot investigate whether the representation of asylum seekers by English-language media outlets to potential non-Chinese audience may also vary based on each outlet's political affiliation.

3. Methods

Moreover, solely focusing on textual content of the media reports means discarding graphical and (if also published online) audio-visual materials accompanied in a given news report, which may contain additional information that are not observed in the texts of news articles. Nevertheless, analysing the textual content of the media reports should still yield a considerable amount of information about the representation of asylum seekers in Hong Kong by media outlets linked to different political camps, since after all a large proportion of the content in news reports are represented textually.

To search for relevant media reports on WiseNews, I will refer to the keywords used by Ng, Choi, et al. (2019) which are commonly used for referring to non-refoulement claimants by the media. These terms can either be directly describing this group of population (e.g. refugees, illegal immigrants) or policies related to asylum seeking (e.g. non-refoulement claim, immigration paper). Table 3.1 lists all the Chinese terms used for searching the relevant news articles on WiseNews and their meanings in English. Additional notes are used for indicating whether some articles were omitted and why. The author also omitted articles only mentioning the keywords briefly throughout the main text which discusses issues not directly related to asylum seekers. As for the keywords “South Asian”, “illegal immigration” and “human snake”, the author only included news entries fetched from these keywords if at least one of the other keywords directly referring to asylum seekers (i.e. non-refoulement claim(ant), torture claim(ant), asylum seeking, refugee and immigration paper) co-appeared in either the title or main text. This is to ensure that the news articles in the dataset are obviously describing non-refoulement claimants in Hong Kong.

Table 3.1: Translation of the keywords used to search for media reports on WiseNews

Original terms in Chinese pinyin	Literal meaning in English	Notes
mian3 qian3fan3 sheng1qing3	Non-refoulement claim(ant)	
ku4xing2 sheng1qing3	Torture claim(ant)	
fei1fa3 ru4jing4, ren2she2	Illegal immigration, human snake (alias of illegal immigrants)	Excluded illegal immigration from mainland China
xun2qiu2 bi4hu4	Asylum seeking	Excluded asylum seekers <i>outside of</i> Hong Kong

3. Methods

Original terms in Chinese pinyin	Literal meaning in English	Notes
nan4min2	Refugee	Excluded refugees <i>outside of</i> Hong Kong and Hong Kong activists seeking refuge overseas
nan2ya4	South Asian	Articles about the South Asian region are excluded.
xing2 jie1 zhi3	Immigration paper	Documents issued by the Immigration Department to non-refoulement claimants.

3.3 Machine learning analysis

3.3.1 Model selection

While the dataset can certainly be analysed via human coding to qualitatively investigate in how the usage of language in reporting on non-refoulement claimants may vary by political camp, due to my motivation of testing the efficacy of machine-learning-based sentiment analysis on non-Chinese-language as well as concern on time and resource intensiveness of manually coding over 500 articles in the dataset, I will instead use supervised machine learning models to test H_1 . In other words, H_1 will be treated as a sentiment analysis problem which classifies the *polarity* of the news articles (i.e. whether they reported on asylum seekers *positively*, *neutrally* or *negatively*). How the independent (both the media outlet's political camp and other controls) and dependent variables are operationalised and coded will be discussed in the next section, and for now I will talk about what kinds of tools and models will be used for the analysis. Nevertheless, it is important to remind that even if machine learning methods can potentially improve the scalability of sentiment analysis on larger datasets, there are also caveats listed by van Atteveldt et al. (2021) such as accidentally identifying spurious relationships between variables or overfitting which can hamper the validity of the analysis. Thus, it is important that I keep these issues in mind while modelling the data and adopt remedies to these problems if necessary.

Starting with the models that will be used for the analysis, I will rely on some commonly used classification models. Candidate models thus include logistic regression, support vector machine (SVM), random forest and gradient boosting. Although van Atteveldt et al. (2021) also mention that deep learning is another popular choice of machine learning model for sentiment analysis tasks, I did not use it due to the relatively small size of the dataset at hand as well as the difficulty

3. Methods

of interpreting the features' impact on the dependent variable directly from deep learning models. I will first run some baseline models with the default parameter settings of the models before selecting one for hyper-parameter tuning based on their performance on their *f1 scores* (which measures the harmonic mean of *precision* and *recall*) on both the test data set (which will be 20% of observations in the original dataset stratified by the dependent variable) and 5-fold cross validation⁴. Compared to metrics like accuracy and the ROC-AUC score, using f1 score as the evaluation metric can better measure a model's performance if there is class imbalance in the dependent variable.

The reason for relying on the f1 score generated from both the testing set and 5-fold cross validation for model selection is because of the inherently small size of the dataset. In the case of having small datasets, using only the test data for model evaluation may risk over- or under-estimating the performance of the model due to the small number of observations only available in the testing data (Brownlee 2020). On the other hand, only using 5-fold cross validation on the whole dataset for training the model cannot have any indication about how the model may perform on predicting data which are completely unseen beforehand to test for its ability to generalisation. Consequently, I decided to take into account of each model's f1 score in both scenarios for a more comprehensive assessment of their performance. The model which overall performs the best will then be used for calculating the *SHAP values* of the features which evaluate how much impact each feature has on the model prediction when the features is at certain values versus at its baseline value (Lundberg and Lee 2017). In essence, the higher the magnitude of a feature's SHAP value, the larger its impact of the model's prediction. And for classification models, positive SHAP value means that an observation is *more* likely to be predicted for belonging to a certain class in the dependent variable, whereas negative SHAP values mean an observation is *less* likely to be predicted for belonging to a certain class.

3.3.2 Data pre-processing

To pre-process the articles into suitable formats as machine learning model inputs, I will also transform the news articles into a term-document matrix which records the frequency of each word (i.e. token) appearing in each article (i.e. document). Instead of simply counting the frequency

⁴Of course, Oriental Daily News contributed to a huge volume of reportage on non-refoulement claimants within the pro-Beijing camp. I will explain why it is not appropriate to re-run the model without including entries from this outlet in the Discussion section below.

3. Methods

of each token in the document corpus, I will use the term-frequency inverse-document-frequency (TF-IDF) matrix transformation which also scales down the importance of tokens that appear in a large proportion of the documents in the corpus since these tokens are likely to be less informative on the content of the documents than tokens that appear in smaller proportions of the documents (Pedregosa et al. 2011). Note that only tokens which appears in all least 2% of the documents in the corpus will be included in the term-document matrix so that I do not include too much noise in the resulting TF-IDF matrix.

Due to the relatively small size of the dataset (with only around 550 articles), I will also adopt non-negative matrix factorisation (NMF) to reduce the dimensionality of the dataset to prevent overfitting. According to Stevens et al. (2012) (p.953), the matrix denoted as H which captures the weight of each topic (as columns) in each document (as rows) of the corpus can help summarise the information of the articles in terms of which topic(s) they primarily focus on. Thus, I will use NMF to transform the TF-IDF matrix into another which summarises the news articles into a certain number of pre-defined latent topics (the exact number will be figured out at the *preprocessing* section in next chapter), and then use this matrix as part of the inputs for the models.

There are also some model-specific pre-processing steps (e.g. standardising numerical features so that their means are centred at 0 and are on the same scale) which will need to be performed accordingly. I will detail them out as necessary at the **Sentiment analysis** section in next chapter.

3.3.3 Implementation

The majority of the implementation from data pre-processing, modelling to validation will be done via the `scikit-learn` package in Python (Pedregosa et al. 2011), whereas the tokenisation of the Chinese-language news articles will be done by the `jieba` package which is designed for natural language processing in Chinese (Junyi 2021). For better tokenisation of the articles, I will also remove stop words, punctuations and hyper-links of the texts and add words related to Hong Kong politics (downloaded from elgarteo (2021)) and asylum seekers into the dictionary. As for the gradient boosting model, the `XGBoost` package (Chen et al. 2015) will be used since it

3. Methods

is one of the most renowned machine learning models in data science, and it also offers a lot of options for tuning the model for better performance.

3.4 Operationalisation

3.4.1 Main independent variable (a.k.a main feature)

Starting from the main independent variable which classifies the associated political camp of the media outlets, I will mainly refer to the literature exploring how Hong Kong’s media industry is intertwined with politics and perform additional research when needed. I consulted the articles by Fong (2017a) and Lee (2018) on the development of post-1997 Hong Kong’s media industry, of which they have documented the co-optation of Hong Kong media outlets by the Chinese Government (e.g. the major shareholders holding positions in such as Chinese People’s Political Consultative Conference). Building on the concept of “allocative control” (p.11) by Lee (2018) about media outlet owners having the ultimate decision on the personnel and resources at the newsroom’s disposal for reporting news which may lead to self-censorship of the frontline journalists, I will code media outlets whose owners are affiliated with the official Chinese and/or Hong Kong Government organisations as belonging to the pro-Beijing camp. As for those media owners with less apparent links with the Chinese or Hong Kong Governments, I will decide the classification according to whether the owners are known to be affiliated with a certain political camp and how a given media outlet is perceived for its political stance. Media outlets without apparent political stances are labelled as “neutral”. Table 3.2 is the resulting classification scheme. *Note that due to the medium of publication of localist news media being overwhelmingly online, none of the outlets in that camp was included in the dataset. Therefore, it would not be possible to test whether the sentiment of reportage on non-refoulement claimants in Hong Kong may be related to localist camp affiliation of media outlets.*

Table 3.2: Classification of the associated political camp of media outlets in Hong Kong included into the sample, source: Fong (2017a); Lee (2018); author’s own research

Associated political camp	Media outlets
Pro-Beijing camp (directly owned by the Liaison Office of the Central People’s Government or the HKSAR Government)	Wen Wei Po, Hong Kong Government News, Hong Kong Commercial Daily, Ta Kung Pao

3. Methods

Associated political camp	Media outlets
Pro-Beijing Camp (owners affiliated with political institutions on mainland China/ HKSAR Government)	Oriental Daily News, Sing Tao Daily, Sing Pao, Headline Daily, Hong Kong Economic Journal, Hong Kong Economic Times, am730, Sky Post, HK01 Newspaper
Pro-Democracy camp	Apple Daily, Kung Kao Po ⁵
Localist camp	N/A
Neutral	Metro Daily, Ming Pao Daily News

Publication of news articles by outlet and political camp

Figure 3.1 shows the number of news articles on non-refoulement claimants published in 2019 by newspaper outlet (left) and political camp (right). Consistent with the study by Ng, Choi, et al. (2019), Oriental Daily News continues to be the media outlet covering the most frequently on asylum seekers with 384 (or 68.94%) articles throughout 2019. By contrast, the second-most frequent publisher *Sing Tao Daily* only had 45 entries (or 8.08%) of the total number of articles published. Each of the other newspaper outlets only constituted to a small portion of news articles about non-refoulement claimants in 2019. Therefore, the issue of asylum seekers in Hong Kong still appeared to be the most salient for Oriental Daily News by 2019, evidenced by its unmatched volume of articles related to this topic vis-a-vis other media outlets.

If we look up the number of news articles on non-refoulement claimants by political camps, the pro-Beijing camp dominated the coverage of asylum seekers in Hong Kong in 2019 largely due to the huge volume of publication by Oriental Daily News. Meanwhile, both neutral and pro-democracy newspaper outlets published similar amounts of articles throughout 2019, and both camps constituted to small proportions of the share of articles during the year. Even if we omitted the sheer volume of articles published by Oriental Daily News, the pro-Beijing media would still have 173 articles published altogether which was still considerably more than the quantity of articles written by neutral and pro-democracy media outlets combined. A question that may arise is that given the comparatively much smaller volume of publication on non-refoulement claimants by neutral and pro-democracy newspaper outlets, is it likely that the articles published by outlets from these two camps will still have potential influence in shaping the narrative on

⁵The decision to put *Kung Kao Po* under the pro-democracy category is made after using the keyword *anti-extradition* (*fan3xiu1li4*) in Chinese to search for relevant news articles on the newspaper's website, which yielded entries that are overall sympathetic to the protesters and disapprove the HKSAR Government's handling of the protest.

3. Methods

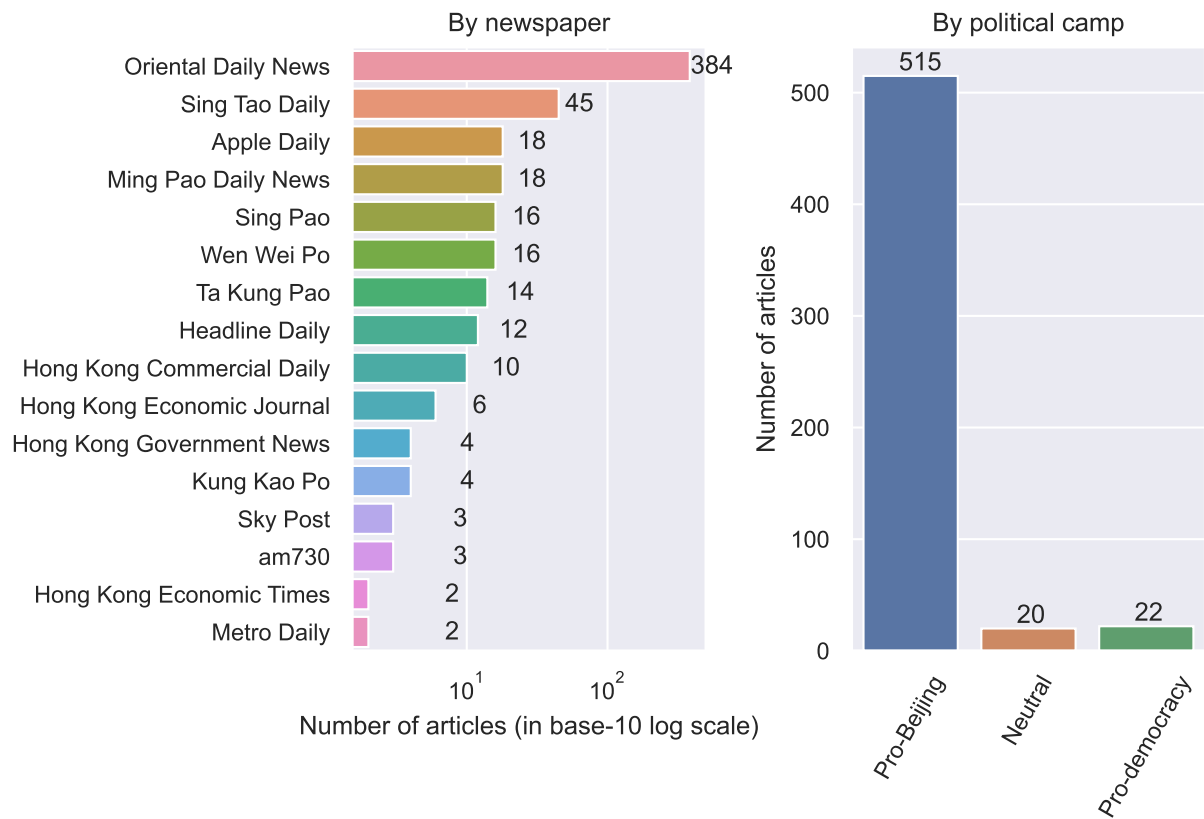


Figure 3.1: News articles on asylum seekers in 2019 by news outlet (left) and political camp (right)

asylum seekers in the broader society? The answer to this question is a rather assuring yes for these two groups of outlets.

Starting with outlets in the pro-democracy camp, *Apple Daily* is one of the most popular paid newspapers in Hong Kong since it occupied over half (54.7%) of the market share in the paid newspaper market and was ranked as the second most popular printed newspaper in Hong Kong by 2019 (Adintime 2021). Apart from being widely circulated in society, *Apple Daily* was also known for its pronounced pro-democracy stance, and it gradually transformed from a pure sensationalist tabloid to flagship newspaper which framed and reported news that were critical to the Beijing and HKSAR Governments since the early 2000s which contributed to its growing popularity, and the forced closure of this outlet was also reported as one of the greatest shocks towards the city's freedom of press by foreign media outlets like the BBC (Tan 2021). As for *Kung Kao Po*, it is a weekly published newspaper founded in 1928 for the Catholic Hong Kong Diocese which often organised religious events to commemorate pro-democracy protests such as

3. Methods

the Umbrella Movement (Po 2019), and given Hong Kong has around 12% of its population identifying as Christians with some of the believers being quite vocal and prominent in pro-democracy activism (Hitchens 2021), the reportage of this newspaper on asylum seekers in Hong Kong can thus be treated as representing the narrative on this issue via the perspective of pro-democracy Christians. Therefore, even if these two pro-democracy outlets published much fewer articles than the pro-Beijing camp outlets on non-refoulement claimants, they may still carry influence over the public's perception towards asylum seekers within the city via framing this issue due to their considerable reader bases.

As for the neutral outlets, it is necessary to include their articles on non-refoulement claimants into the dataset because they also occupy an important role in Hong Kong's media industry by providing objective media coverage which avoids explicit political allegiance with a particular camp. Despite the controversial change of chief editorship in 2014, Lee (2018) argues that Ming Pao has established itself as one of the "widely circulated, influential and credible" (p.12) outlets in Hong Kong since the 1980s, and its owner *Tiong Hiew King* has rather emphasised that Ming Pao report news objectively similar to the Chinese newspapers under his ownership in Malaysia. Likewise, being the first free printed newspaper available in Hong Kong from 2002 to 2019 (which then became an online newspaper since 2020) as the localised edition of *Metro International*, including Metro Daily as one of the outlets covered in the dataset can also help investigate whether and how local editions of international media may pay attention to the issue of asylum seekers in Hong Kong (as it turned out Metro Daily hardly covered this issue during its last year of publication as a printed newspaper).

In short, even if the number of articles on non-refoulement claimants published by the neutral and pro-democracy outlets included in the dataset was small, their reportage should still be able to reach out to a considerable amount of audience to potentially influence the latter's perception towards asylum seekers because of the outlets' wide circulation as well as their status as some of the leading media outlets in their respective political stance.

3.4.2 Dependent variable (a.k.a target variable)

As mentioned before, the dependent variable of H_1 will be the sentiment of the news articles towards asylum seekers in the polarities of *positive*, *neutral* and *negative*. But how exactly should

3. Methods

these three polarities mean in this context? Again referring to the literature can help devise a sound coding scheme. Drawing McKay et al. (2011), Cooper et al. (2017) and Ng, Choi, et al. (2019), *positive* articles usually acclaim the person or event in concern, describe personal stories, report expert opinions without derogatory terms or depict refugees favourably, *neutral* articles simply report an incident without an overt sentiment or opinion towards particular stakeholders and with little interpretation by the journalists, and *negative* articles frame asylum seekers either as bringing problems to the host society or use inaccurate terms to delegitimise this group of population. I will follow these coding criteria for the polarities of the news articles in the dataset since these definitions suit the purpose of this study. The numerical representation of the polarities in the dataset will be as follows, namely, positive articles will be coded as 2, neutral articles will be coded as 1, and negative articles will be coded as 0.

For the context of Hong Kong about the “incorrect” terms and information used for describing asylum seekers, I will refer to expert opinions about whether descriptions of asylum seekers in Hong Kong are founded⁶. For example, Ng, Choi, et al. (2019) point out that the term “fake/bogus refugees” (*jia3 nan4min2*) should be considered as both derogatory and inaccurate because those applying for non-refoulement claims are not automatically qualified as refugees while awaiting the final decisions of their applications. *Therefore, the polarity of news articles in which the news media directly refers non-refoulement claimants as “fake refugees” in its reportage will be coded as negative, but the report should be coded at least as neutral instead if the news outlet is merely quoting someone who describes non-refoulement claimants as “fake refugees”.* Furthermore, Ng (2020) mentions that non-refoulement claimants are forced to be illegal immigrants by the policy design of the HKSAR Government so that this group of population can be excluded from social rights that legal residents enjoy. News articles equating non-refoulement claimants as illegal immigrants will thus also be coded as negative since such reportage is justifying the exclusion of non-refoulement claimants by resorting to their illegality that had been forced upon by the HKSAR Government.

⁶A detailed discussion of the depiction of asylum seekers in Hong Kong can be found in Ng (2020) and Ng, Choi, et al. (2019).

3.4.3 Control variables (a.k.a features)

I will also incorporate some control variables to see if the association between political camps and polarities of the news articles still remains significant afterwards. For starters, the time when the articles were published will be included as a control, though I will bin the twelve months into four evenly split quarters (*Quarter*) to avoid this category from having too high cardinality which may hamper the performance of tree-based models⁷, with Q1 covering from January to March, Q2 covering from April to June, Q3 covering from July to September, and Q4 covering from October to December. As 2019 was also the year when one of the most large-scale and long-lasting pro-democracy protests broke out, it would be interesting to see if the polarities of the news articles might vary according to in which Quarter it was published, and one could argue that the anti-extradition law witnessed some of the most intense conflicts during Q3 and Q4.

Moreover, the presence of racial labels in the news articles will be included as a control since it is possible that more overtly racialised discourse may be associated with more negative sentiments towards asylum seekers in general according to the literature. This variable will be named *Racial_label* in the dataset and will be binary, with 1 meaning there is presence of racial labels in the main text and 0 otherwise. Racial labels referring to both regional (e.g. South Asia, Africa) and national (e.g. the nationalities included in figure 2.2) will be counted. Lastly, the metadata of the news articles, namely the character counts of titles and main texts, will also be added to see if the lengths of the articles may be associated with their sentiment towards asylum seekers.

⁷A detailed discussion of the potential problems posed by highly cardinal categorical features on tree-based models can be found in Ler (2021), but the main issue is that the higher the cardinality of a categorical variable, the more splits will be needed by tree-based models which may then considerably increase the computational costs of the models.

Chapter 4

Results

After the data collection process, there were in total 557 articles published in 2019 by 16 newspapers which reported on non-refoulement claimants residing in Hong Kong. I will first preliminarily explore the variables in the data set before moving onto running sentiment analysis with machine learning models to find out whether the political camp of media outlets is associated with the polarity of the news articles towards asylum seekers. Interpretation of the model's results will be provided afterwards.

4.1 Exploratory data analysis (EDA)

4.1.1 Number of news articles by month

Figure 4.1 shows how the number of news articles on non-refoulement claimants varied by month in 2019. It appears that coincidentally, there were fewer articles about asylum seekers published between August and November when some of the most intense clashes (notably the siege of the Hong Kong Polytechnic University in November 2019) took place, and only until December did the amount of published articles about non-refoulement claimants rebound to the level in July. As for the first half of 2019, the amount of articles published by month fluctuated quite a bit as seen from the zig-zag pattern of the line graph on this period.

4.1.2 Polarities of the news articles

According to table 4.1, the polarity of the news articles about asylum seekers in Hong Kong in 2019 tilted towards negative, since only around 4.3% and 23.5% of articles respectively depicted asylum seekers positively and neutrally. The fact that the sentiment of the news articles in 2019 was skewed towards negativity implies that I will need to take class imbalance into account

4. Results

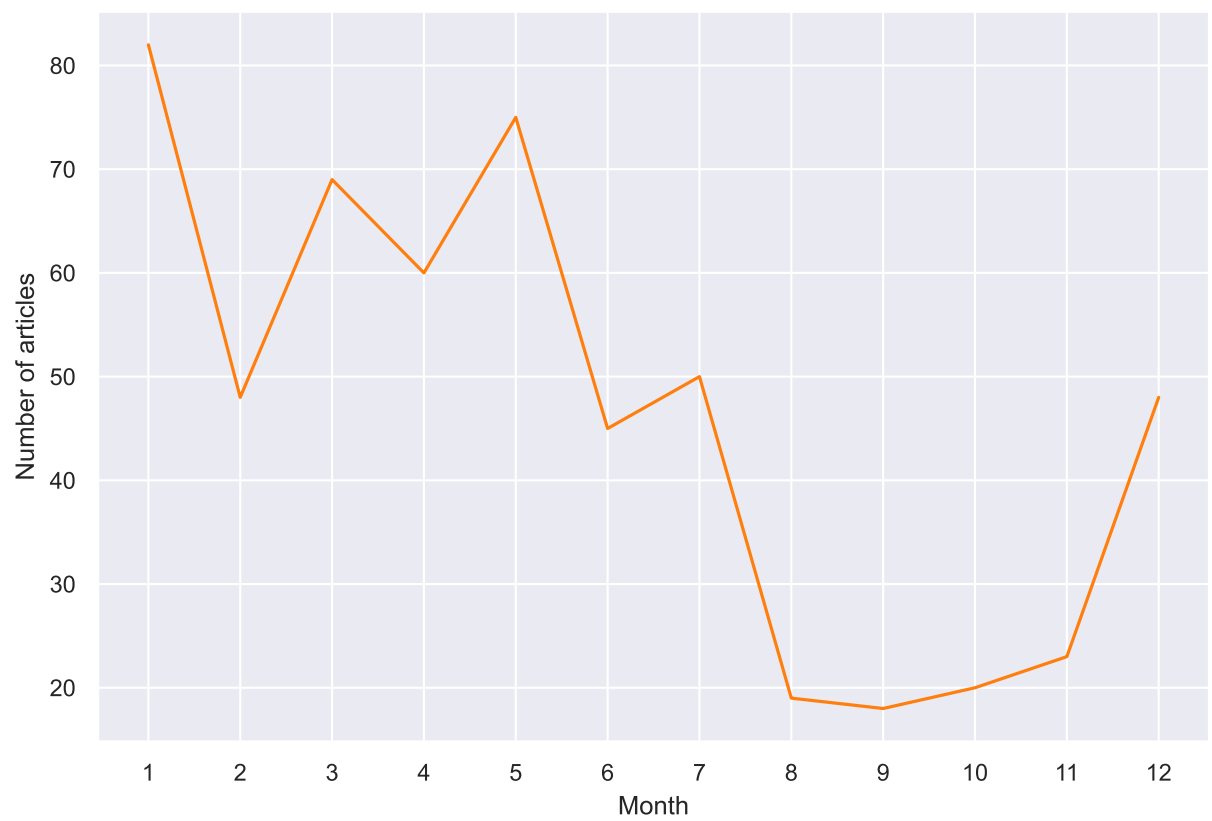


Figure 4.1: Temporal patterns of the publication of news articles about asylum seekers in Hong Kong in 2019

4. Results

Table 4.1: Polarities of the news articles on asylum seekers in Hong Kong in 2019

	Negative	Neutral	Positive	All
Neutral	0	12	8	20
Pro-Beijing	402	108	5	515
Pro-democracy	0	11	11	22
All	402	131	24	557

for modelling later. Political-camp-wise, pro-Beijing media outlets had over 70% of its articles depicting asylum seekers in Hong Kong in negative lights, whereas neutral and pro-democracy media outlets had their reportage evenly spread between neutral and positive articles (albeit they altogether constituted to only a small proportion of the total number of articles in 2019). While H_1 shall be tested formally with machine learning models after including other control variables later, preliminary evidence suggests that the polarities of the news articles vary with the political camp that the outlets belong to.

4.1.3 Presence of racial labels

Given the majority of asylum seekers in Hong Kong being non-ethnic Chinese, it will also be worth glimpsing whether the presence of racial labels for describing asylum seekers is associated with the sentiment of the news articles. Judging from figure 4.2 preliminarily, however, it appears that the patterns of the polarities are quite similar whether news articles contain racial labels or not, namely, most of the articles framed non-refoulement claimants negatively, some reported on events about this group of population neutrally, and only a small amount of articles were favourable towards asylum seekers residing in the city. Machine learning models can test whether there is any association between racial labels and the polarity of news articles more formally later.

4.1.4 Character lengths of news articles and titles

Lastly, let's look at the distribution of the character lengths of the titles and main texts of the news articles. According to figure 4.3 and table 4.2, it appears that both the title and main text lengths have right-skewed distributions. In other words, while most of the news articles on asylum seekers in Hong Kong in 2019 had relatively short titles and/or main texts, a few of them were considerably more verbose than the rest.

4. Results

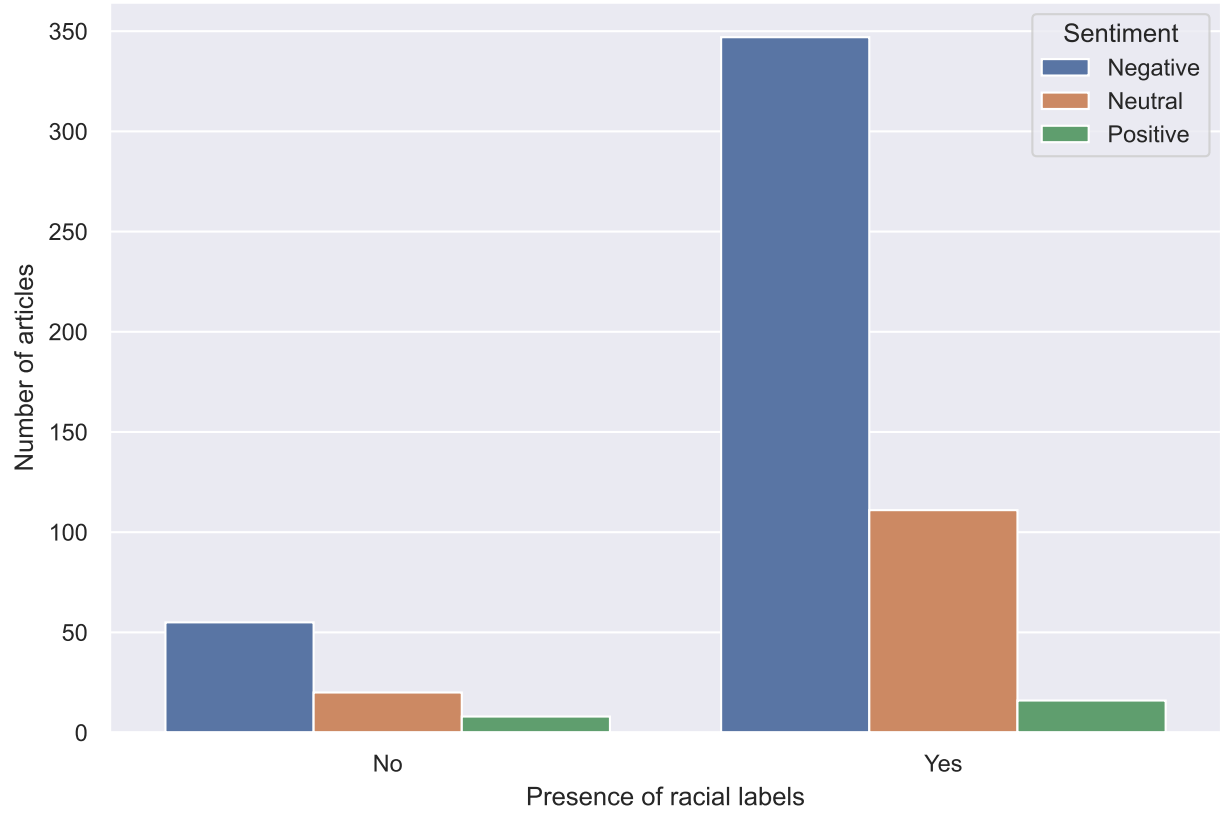


Figure 4.2: Presence of racial labels in the news articles by sentiment

Table 4.2: Summary statistics of the word counts of the news articles' titles and main texts

	Title	Raw main text
count	557.000000	557.0000
mean	15.965889	683.9264
std	5.993154	453.9607
min	4.000000	80.0000
25%	12.000000	404.0000
50%	15.000000	581.0000
75%	18.000000	893.0000
max	48.000000	4715.0000

4. Results

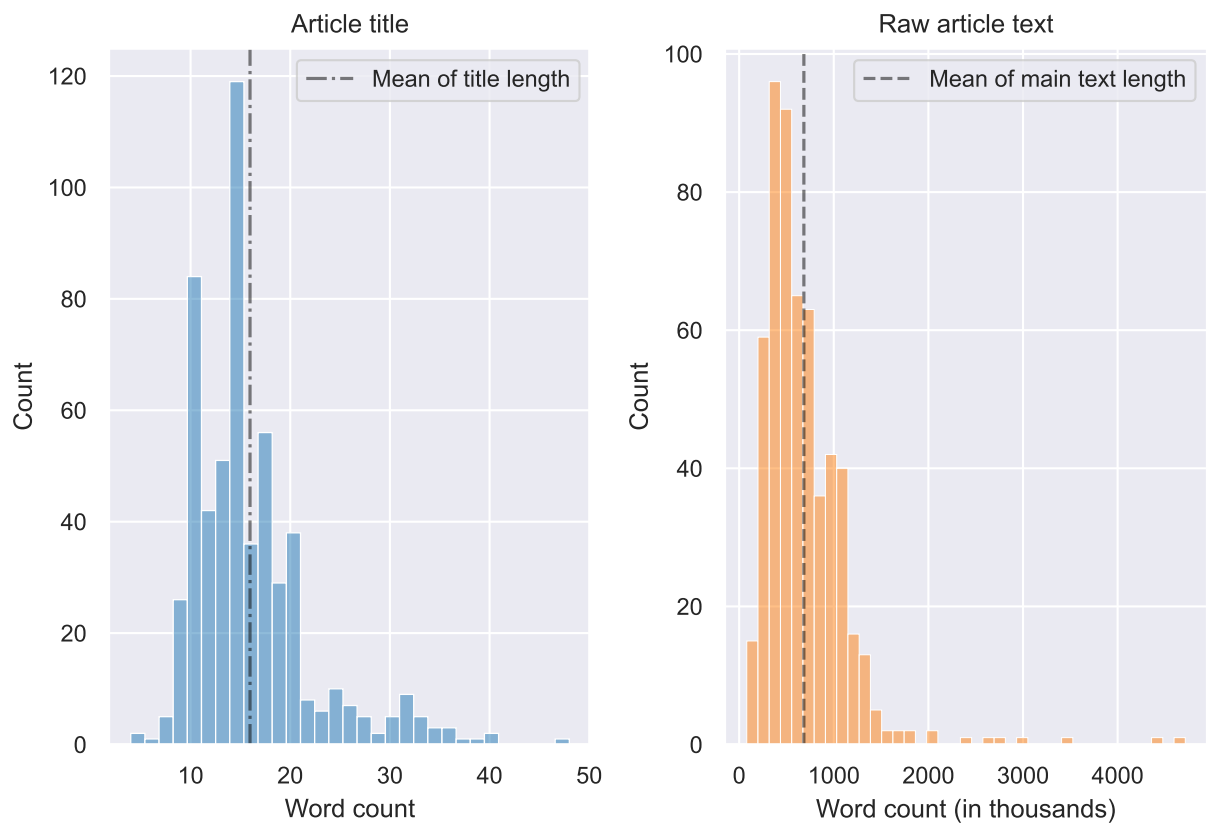


Figure 4.3: Distributions of the word counts of the articles' titles (left) and main texts (right)

4. Results

4.2 Sentiment analysis

4.2.1 Pre-processing

After making sense of the dataset with EDA, it is time to build the sentiment analysis model to see whether the political affiliation of news media outlets is associated with the polarities of the news articles after controlling for other variables. But first there are some pre-processing steps to be done so that the data are transformed into suitable formats as inputs for machine learning models. For starters, columns of the metadata should be excluded for being the inputs of the models. Note that I have also removed the `Newspaper` column since H_1 is more interested in whether newspaper outlets of the pro-Beijing camp *as a whole* may hold more negative attitudes towards asylum seekers in Hong Kong vis-a-vis media outlets with other political stances. The removed metadata columns are: `Index`, `Date`, `Category`, `Page_number` and `Newspaper`.

For the categorical features, they will need to be transformed via one-hot encoding, meaning that each of them will be transformed into n (i.e. the number of the original distinct values in each categorical feature) new columns, and the presence of a category will be marked with 1 in the respective column for an observation and 0 for other columns. Meanwhile, it would also be better to standardise the numerical features (i.e. other than `Political_camp` and `Quarter`) by centering their means at 0 for better model convergence, but the standardiser should only be fitted on the *training set* after splitting the data into the training and testing sets in order to avoid data leakage (the same is also true for creating the TF-IDF matrix). As mentioned before, 20% of the observations in the dataset will be split into the test data for model validation later.

```
from sklearn.model_selection import train_test_split

X = news_df.drop(columns="Sentiment")
y = news_df.Sentiment

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y, random_state=1)
```

The next step is to transform both the titles and main texts of the articles into a TF-IDF term-document matrix. Apart from joining the `Title` and `Text` columns together as the complete `Article`, I will also add additional words into the dictionary and remove stop words as well as punctuation for better tokenisation so that the NMF model can better discover the latent topics.

4. Results

```
# Train set
X_train["Article"] = X_train.Title.str.cat(news_df.Text, sep=" ")
X_train.drop(columns=["Text", "Title"], inplace=True)

# Test set
X_test["Article"] = X_test.Title.str.cat(news_df.Text, sep=" ")
X_test.drop(columns=["Text", "Title"], inplace=True)
```

```
def read_text(path):
    with open(path, 'r', encoding='utf-8') as file:
        text = file.readlines()
        text = [word.replace('\n', ' ') for word in text]
    return text
```

```
hk_politics_words = read_text('Coding/HKPolDict-master/merged.txt') # Words related to Hong Kong poli
asylum_seeker_words = read_text('Coding/Asylum_seeker_words.txt') # Words related to asylum seekers i
for word in chain(hk_politics_words, asylum_seeker_words):
    jieba.add_word(word)
```

```
## Building prefix dict from the default dictionary ...
## Loading model from cache C:\Users\kenji\AppData\Local\Temp\jieba.cache
## Loading model cost 0.477 seconds.
## Prefix dict has been built successfully.
```

To avoid data leakage as mentioned before, I will only fit the `TfidfVectorizer` and the NMF models on the train set (i.e. `X_train`) and then use the fitted instances to transform both the train and test data. I set the number of latent topics (`n_components`) as 10 for the NMF model, and this is decided based on figure 4.4 which plots the reconstruction error measuring the difference of the values between the original TF-IDF matrix and the reconstructed version after NMF. Although there are certainly other valid choices of the number of latent topics to be discovered by NMF, 10

4. Results

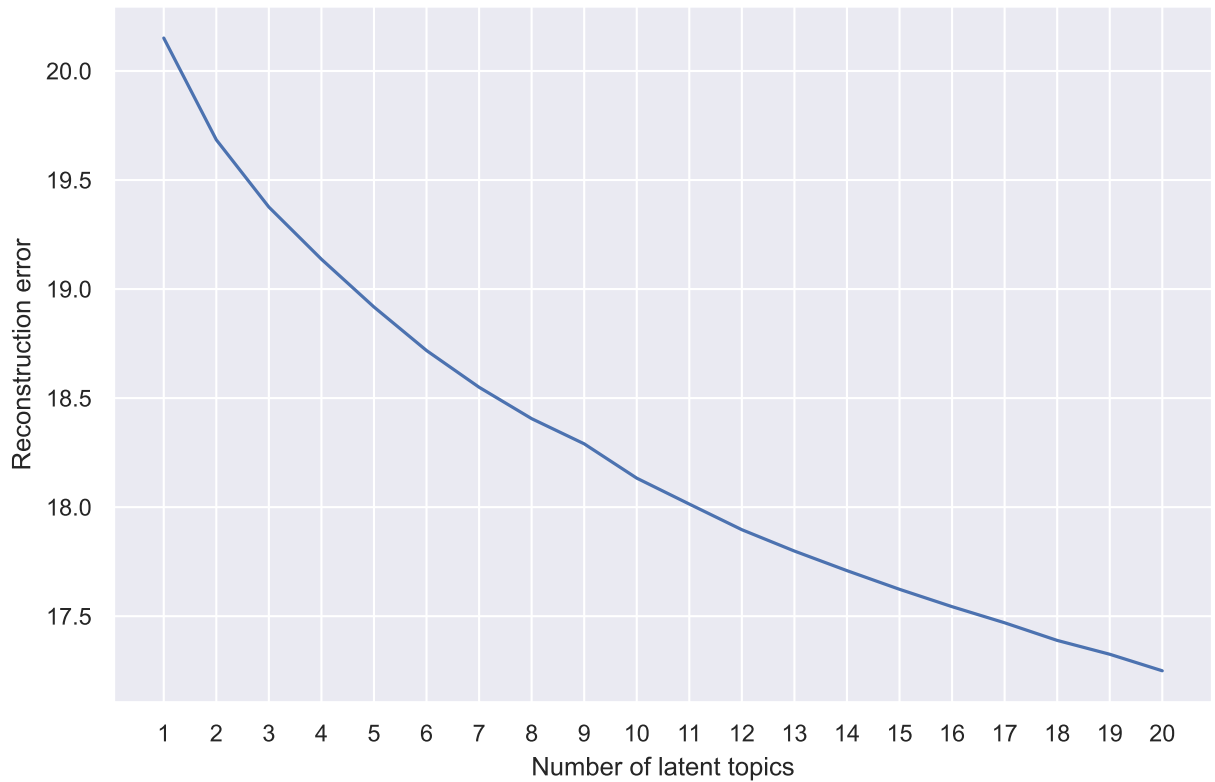


Figure 4.4: Elbow plot of the reconstruction error of NMF as a function of the number of pre-specified latent topics

appears to be a reasonable choice as a compromise between finding out a wide variety of topics in the corpus and not fitting too much into the noise of the data.

In order to make the latent topics generated by the NMF model be named more intuitive, I will inspect the 30 most prominent words of each latent topic (which are shown in figure 4.5) and then summarise each topic. Overall, the ten topics generated by NMF are more or less semantically coherent. Finally, I will transform the validation set’s articles with the fitted instance of the TF-IDF and NMF models on the training set.

The ten latent topics (from topics 1 to 10) that were discovered by the NMF model can be named as: *crimes*, *non-refoulement legal procedure*, *illegal labours*, *illegal gambling*, *drugs*, *illegal immigration*, *murder*, *robbery*, *South Asian settlements* and *problem and solution*. The *problem and solution* topic refers to the coverage on whether and how non-refoulement claimants should be treated as a social problem and how this should be “solved” accordingly, and it should be noted that the scope of this topic is broader than the rest, meaning that there is likely to be

4. Results

Topic 1 (crimes):
南亞 李主 漢 現場 兵團 一名 受傷 刑事 報警 調查隊 警方 犯案 調查 附近 案件 凌晨 兩名 男子 事件簿 送院 警區 治理 追緝 逃 初步 接報 進 求助 交由
Topic 2 (non-refoulement legal procedure):
聲請 處理 司法覆核 個案 免遣返聲請 提出 上訴 立法會 宗 法律 入境處 政府 審核 程序 假難民 保安局 遣返 修例 修訂 建議 議員 相關 問題 年度 當局 提交 考慮 指 申請 司法機構
Topic 3 (illegal labours):
入境處 非法 黑工 勞工 監禁 被判 工作 僱傭 聘用 介乎 被捕 年齡 行動 定罪 最高 僱主 一經 離境 十五 罰款 工行 月 餐廳 擔保書 拘捕 一名 入境者 涉嫌 女子 身份
Topic 4 (illegal gambling):
遊戲機 賭檔 中心 黑幫 賭博 仔 賭博場 行動 賭客 警方 持牌 非法 包括 魚機 遊戲 經營 釣 機 戲機 積分 被捕 釣魚機 集團 南亞 現金 三合 機及 人士 舖 萬元
Topic 5 (drugs):
毒品 檢獲 約 販毒 海關 大麻 單位 警方 行動 懷疑 男子 分銷 可卡因 被捕 市值 拘捕 鈔 涉嫌 調查 冰毒 藏毒隊 搜查 探員 追查 來源 特別 南亞 公斤 製 報
Topic 6 (illegal immigration):
人蛇 水管 偷渡 蛇頭 偷渡客 來港 南亞 搶灘 非法入境 西貢 截獲 警方 一艘 巴基斯坦 船 集團 屯門 八名 可疑 發現 水域 深圳 安排 拘捕 船上 窗 西分區 昨晨 籍 海里
Topic 7 (murder):
死者 房 罰 妻子 男子 殺 警方 單位 非洲 行街紙 死亡 同鄉 傷痕 住客 殺案 謀 謀殺 調查 印度 印度籍 發現 爭執 土瓜灣 海防道 九龍城 被捕 昏迷 床上 被告
Topic 8 (robbery):
事主 背囊 賊 搶 休班 匪徒 黃埔 兩名 換店 紅磡 刀 警長 一名 現金 巨款 找 男子 萬元 匪 制服 牛肉 逃 截劫 疑匪 劫 劫案 南亞 合力 疑犯 德安街
Topic 9 (South Asian settlements):
南亞 村 拆 木板 新村 橋底 屋 清 露宿 村民 政府 通州街 深水埗 居民 區議員 公園 部門 天橋底 衛生 街坊 搭建 渡船街 雜物 油麻地 治安 更 假難民 垃圾 吸毒
Topic 10 (problem and solution):
港府 假難民 香港 政府 說 難民 問題 社會 市民 治安 南亞 做 派 更 政策 港人 沒 政治 港 退出 長 毒瘤 禁閉營 移民 加拿大 解決 法治 錢 美國 億元

Figure 4.5: Word list of each of the topics generated by the NMF model, note: the importance of each word in each topic is arranged in descending order from left to right

more ways of how this topic can be covered by newspapers. The resulting NMF matrix will be in the shape of 557 rows with 10 columns, meaning that each news article will be on its own row in which the 10 columns will indicate the weights of the ten latent topics discovered by the NMF model. The higher the weight of a topic, the more emphasised this topic is in an article, and multiple topics may be co-mentioned within a news articles albeit with different weights.

4.2.2 Training the model

After the above pre-processing steps, it is time to train a machine learning model that can adequately predicts the relations between the features and the sentiment of the articles before finding out the importance of the political camp as the main independent variable. To facilitate the decision of which model to use, I will first run some baseline models with the default hyper-parameters, except that I have adjusted the weights of each class in the dependent variable due to the class imbalance problem mentioned earlier and also set the `early_stopping_rounds` argument to 5 to prevent overfitting while training the XGBoost model¹. Moreover, tree-based models (i.e. random forest and XGBoost) do not necessarily need to have the numerical features standardised, and thus only the categorical columns need to be one-hot encoded. The baseline models will be compared based on their performance on the *macro* average f1 score (which is simply the unweighted average of per-class f1 scores, Pedregosa et al. (2011)) on the testing set

¹For the complete documentation of the default parameters of the models used in this thesis, refer to the websites of scikit-learn and XGBoost Documentation — xgboost 1.5.1 documentation.

4. Results

Table 4.3: F1 scores on 5-fold cross validation and test data of the baseline models

	5-fold cv f1 score	Test data f1 score
Logistic regression	0.6909535	0.8109777
SVM	0.6946997	0.8032258
Random forest	0.7936016	0.6848406
XGBoost classifier	0.7479415	0.7988271

Table 4.4: F1 scores on 5-fold cross validation and test data of the baseline and tuned XGBoost models

	5-fold cv f1 score	Test data f1 score
Baseline	0.7479	0.7988
Tuned	0.7616	0.8814

and 5-fold cross validation. I chose the macro average f1 score because there is no apparent reason for treating the prediction of one polarity of the news articles to be more important than others.

Table 4.3 contains the performance of the macro average f1 scores of the four baseline models on both the 5-fold cross validation and test data. It seems that the two tree-based models perform better than the two linear classifiers (i.e. logistic regression and SVM) in the 5-fold cross validation. Eventually, I decided to go forward with the XGBoost model because the random forest model's f1 score on the test data is more than 0.1 less than that on 5-fold cross validation which is indicative of over-fitting, whereas this was not the case for the XGBoost baseline model. Therefore, I will proceed with XGBoost for hyper-parameter tuning to see if the model's performance can be improved².

The comparison of the baseline and tuned model's performance in f1 score can be found in table 4.4³. As expected, the tuned XGBoost model performs better on both 5-fold cross validation and the test data (although the magnitude of the increase of the f1 score on the test data should be taken with a grain of salt because of the small number of observations available in the dataset as mentioned before). Table 4.5 further breaks down the f1 score of the tuned XGBoost model on the test data. I will thus calculate the SHAP values of the features with the tuned XGBoost model given its better overall performance.

²The code for tuning the model can be found in the appendix.

³Hyper-parameters of the tuned XGBoost model used here are: {'objective': 'multi:softprob', 'use_label_encoder': False, 'base_score': 0.5, 'booster': 'gbtree', 'colsample_bylevel': 1, 'colsample_bynode': 1, 'colsample_bytree': 0.6000000000000001, 'enable_categorical': False, 'gamma': 0.5, 'gpu_id': -1, 'importance_type': None, 'interaction_constraints': '', 'learning_rate': 0.37, 'max_delta_step': 0, 'max_depth': 5, 'min_child_weight': 3.0, 'missing': nan, 'monotone_constraints': '()', 'n_estimators': 40, 'n_jobs': 16, 'num_parallel_tree': 1, 'predictor': 'auto', 'random_state': 1, 'reg_alpha': 0, 'reg_lambda': 12.0, 'scale_pos_weight': None, 'subsample': 0.7000000000000001, 'tree_method': 'exact', 'validate_parameters': 1, 'verbosity': None, 'eval_metric': 'mlogloss'}. To reproduce this model, simply create an XGBClassifier instance, copy the above hyperparameters as a dictionary and then use the .set_params(**dict) method on the XGBClassifier instance.

4. Results

Table 4.5: F1 score breakdown of the tuned XGBoost model on the test data

	precision	recall	f1-score	support
0	0.9136	0.9136	0.9136	81.000
1	0.7308	0.7308	0.7308	26.000
2	1.0000	1.0000	1.0000	5.000
accuracy	0.8750	0.8750	0.8750	0.875
macro avg	0.8814	0.8814	0.8814	112.000
weighted avg	0.8750	0.8750	0.8750	112.000

4.3 Results: was pro-Beijing media outlets more likely to negatively portray asylum seekers in Hong Kong in 2019?

This section will report the findings of how important each feature contributed to the model's prediction of the sentiments of the news articles on asylum seekers in Hong Kong by media outlets in 2019. Apart from reporting on whether affiliation with the pro-Beijing camp of media outlets is an influential feature for predicting each of the polarities, I will also mention other worth-noting findings. The data points used for constructing the SHAP values are from the training set (`X_train_final`). According to the documentation of the `TreeExplainer` class used for generating SHAP values of tree-based ensemble models ([Shap.TreeExplainer SHAP Latest Documentation n.d.](#)), in the case of classification tasks, the model output explained by SHAP values with regard to the features is the *log odds ratio*. Consequently, positive SHAP values (i.e. increasing log odds ratio) mean that an observation is more likely to belong to a certain class, whereas negative SHAP values (i.e. decreasing log odds ratio) implies that an observation is less likely to belong to a certain class.

4.3.1 Overall magnitudes of SHAP values for each feature

With the trained model at hand, we can now answer whether H_1 is supported by the model's results using SHAP values. According to figure 4.6, we can see that on the level of the whole model, whether a media belongs to the pro-Beijing camp or not (`Political_camp_Pro-Beijing`) is the third most important features in predicting the sentiment of a news article, and its magnitude of SHAP values in affecting the model's output is only lower than those of two news article topics about asylum seekers (i.e., `Problem and solution` and `Murder`) while being slightly higher than that of articles on the topic of Crimes. Furthermore, within the bar of the SHAP values of the `Political_camp_Pro-Beijing` feature, we can see that pro-Beijing affiliation of media outlets is

4. Results

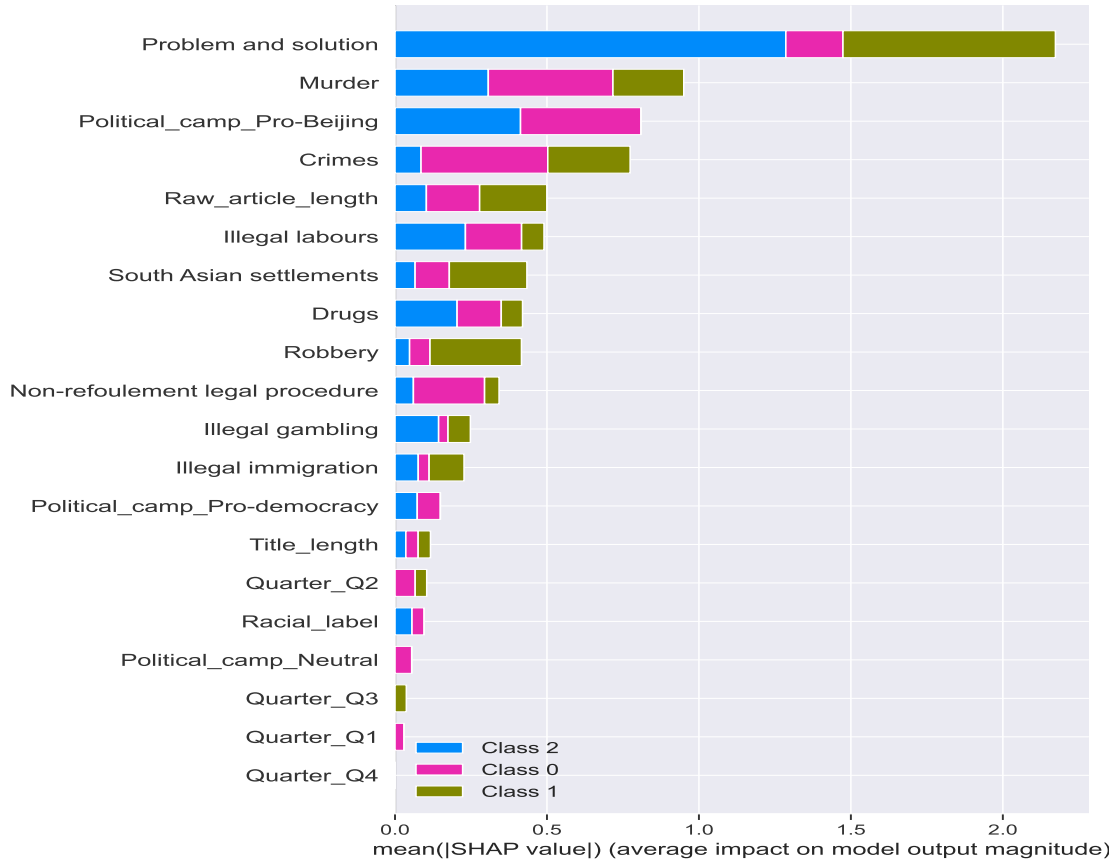


Figure 4.6: Magnitudes of SHAP values of each feature for all classes

considerably more informative for predicting whether an article has positive polarity (Class 2) or not and has negative polarity (Class 0) or not, but not so much while predicting if an article has a neutral polarity (Class 1) or not. Therefore, it seems that pro-Beijing media outlets do tend to differ their sentiments on the reportage of asylum seekers in Hong Kong compared to other non-pro-Beijing outlets.

It should also be noted that contrary to expectation, `Racial_label` is not an informative feature in predicting the sentiment of the articles since its magnitude of SHAP values is at the bottom five out of all features. One potential reason may be that the connotation between non-ethnic Chineseness and non-refoulement claimants is quite salient that these two labels are often used together by newspaper outlets no matter the sentiment or attitudes towards asylum seekers in Hong Kong. Furthermore, the sentiments of the news articles do not seem to correlate with in which quarter they were published, and so is the length of the news articles *titles*.

Nevertheless, figure 4.6 does not really show the *direction* of the SHAP values of each feature

4. Results

in affecting the model's output. We can therefore use the beeswarm summary plots from the `shap` package which also shows the directions of SHAP values for predicting each class in the dependent variable as each feature's value changes. The importance of the features is arranged in descending order on the y-axis from top (the most important) to bottom (the least important). In a SHAP value beeswarm summary plot, dots in black mean the value of a feature is high (or present in case of a binary feature, e.g. one-hot-encoded columns), whereas those in white mean the value of a feature is low (or absent in the case of a binary feature). Moreover, dependence plots which zooms in the relationship between the feature values (on the x-axis) and SHAP values (on the y-axis) of the eight most important features will be provided.

A little note on how to interpret the findings from the beeswarm summary and dependence plots. In essence, each plot will show the SHAP values of each feature to assess whether a feature's value away from its baseline will increase or decrease the model's prediction of the log odds ratio that an observation belongs to a certain class. For instance, if feature A's SHAP value for class 0 increases as its value becomes larger, then this means the model will predict that an observation is more likely to belong to class 0 as feature A's value becomes larger.

4.3.2 SHAP values of predicting negative news articles

According to figure 4.7, pro-Beijing affiliation (`Political_camp_Pro-Beijing`) is the third most important feature in predicting whether a news article reports on asylum seekers in Hong Kong negatively or not. Consistent with the expectation in H_1 , articles published by pro-Beijing media newspapers are more likely to report on asylum seekers negatively than those by outlets with different political orientations. Conversely, albeit with less magnitudes in SHAP values, media outlets with neutral (`Political_camp-Neutral`) or pro-Democracy (`Political_camp_Pro-democracy`) stances are less likely to publish negative articles on non-refoulement claimants. When it comes to negative articles, therefore, political affiliations of media outlets in Hong Kong do matter, meaning that pro-Beijing camp newspapers are more likely to depict non-refoulement claimants negatively than their counterparts holding other political stances⁴.

Figures 4.8 and 4.9 further zoom into the SHAP values of the eight most important features for predicting the negative polarity of the news articles. Some interesting patterns are observed

⁴Of course, Oriental Daily News contributed to a huge volume of reportage on non-refoulement claimants within the pro-Beijing camp. I will explain why it is not appropriate to re-run the model without including entries from this outlet in the Discussion section below.

4. Results

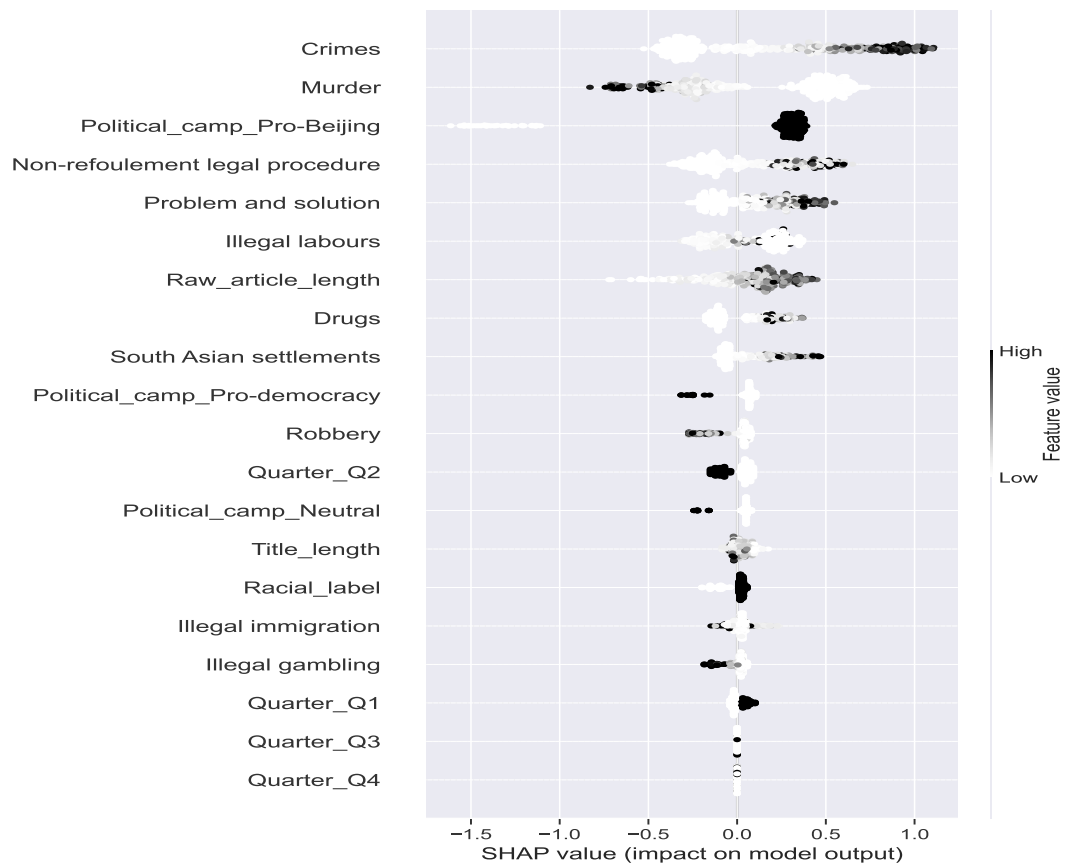


Figure 4.7: The SHAP values of the features in the prediction of whether an article has a negative polarity

4. Results

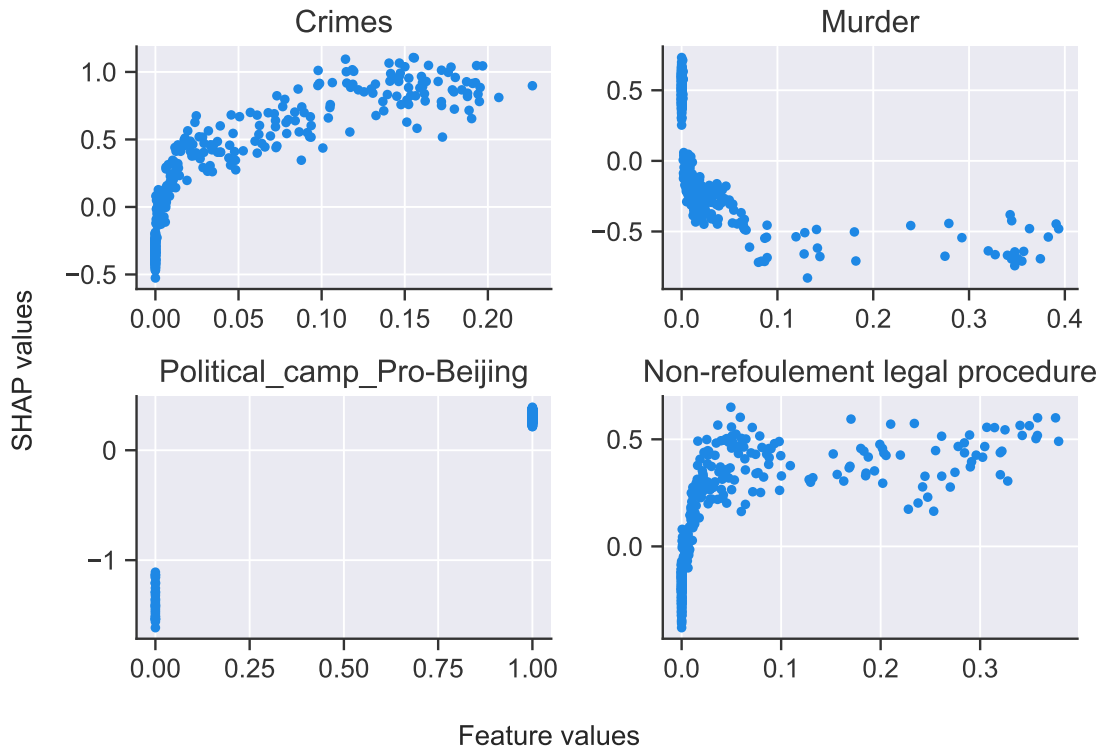


Figure 4.8: Dependence plot of the eight most important features for predicting negative polarity of news articles

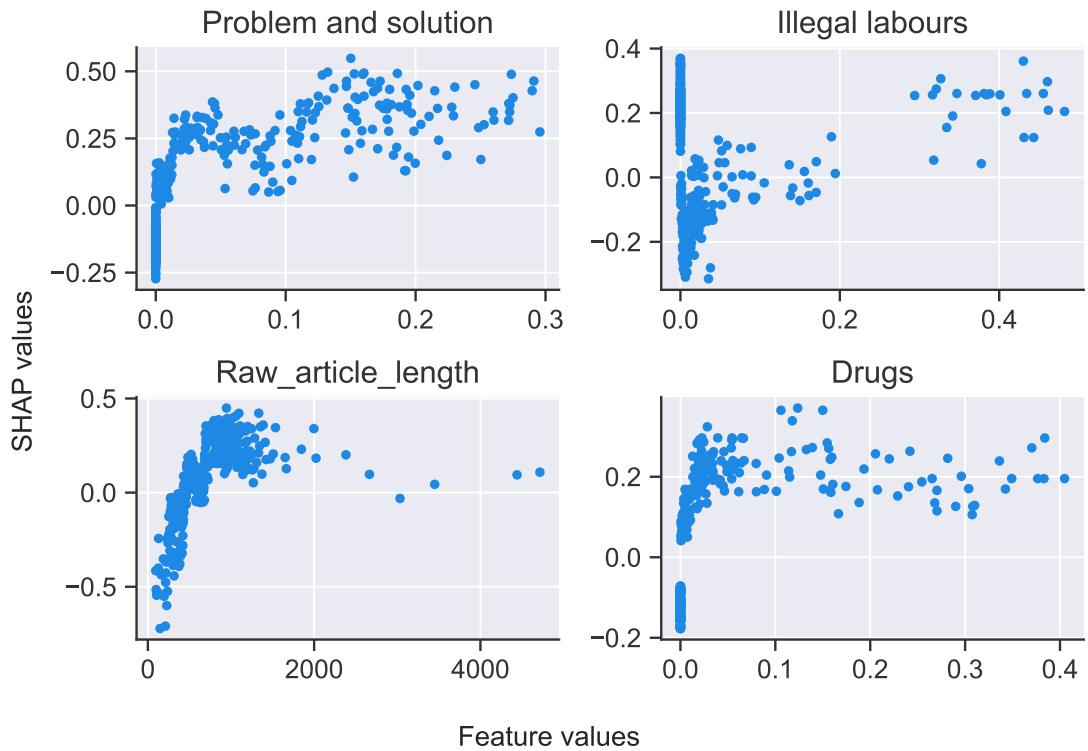


Figure 4.9: Dependence plot of the eight most important features for predicting negative polarity of news articles

4. Results

here. Firstly, articles which are more related to crimes (`Crimes`), the legal procedures of non-refoulement claims (`Non-refoulement legal procedure`) and how non-refoulement claimants may be a problem and the according solutions (`problem and solution`) are more likely to report on asylum seekers negatively. Clues of the potential causes of observing these correlations can be found in the significant topic words associated with each topic shown in figure 4.5. For instance, both the topics of `Non-refoulement legal procedure` and `problem and solution` contain the derogatory term “fake refugee”, whereas the topic `Crimes` found the noun phrase “South Asian army” (*nan2ya4 bing1tuan2*) as two of the topic words which together attempt to conflate non-refoulement claimants with the image of them being South Asians coming to Hong Kong en masse for committing crimes.

Interestingly, more verbose news articles (i.e. having longer `Raw_article_length`) are also predicted to be more likely for holding negative opinion towards non-refoulement claimants. Lastly, a more pronounced presence of the topics of `Illegal labours` and `Drugs` in a news article appear to be more likely to have a negative polarity against non-refoulement claimants as shown by their SHAP values. Meanwhile, articles focusing on the theme of `Murder` are predicted to be *less likely* for carrying a negative polarity.

4.3.3 SHAP values of predicting neutral news articles

Figure 4.10 shows the SHAP values of each feature in contributing to the model’s prediction of whether a news articles is likely to have neutral polarity towards non-refoulement claimants. Contrary to the case of the prediction of negative news articles, the political affiliation of newspaper outlets do not contribute substantively to the model’s output, as all three political orientations of newspaper outlets covered in this thesis are located at the bottom five positions on the y-axis, and their SHAP values are essentially zero.

Rather, some topics of the news articles are more informative in predicting whether news articles are only reporting incidents without much interpretation by the journalists. We can look into this further in figures 4.11 and 4.12. For instance, articles more related to `problem and solution`, `Crimes`, `South Asian settlements` or `Illegal labours` (albeit with smaller magnitude) are less likely to be written neutrally by newspaper outlets covered in the dataset. In other words, these topics are more likely to be loaded with particular sentiments when they were reported by

4. Results

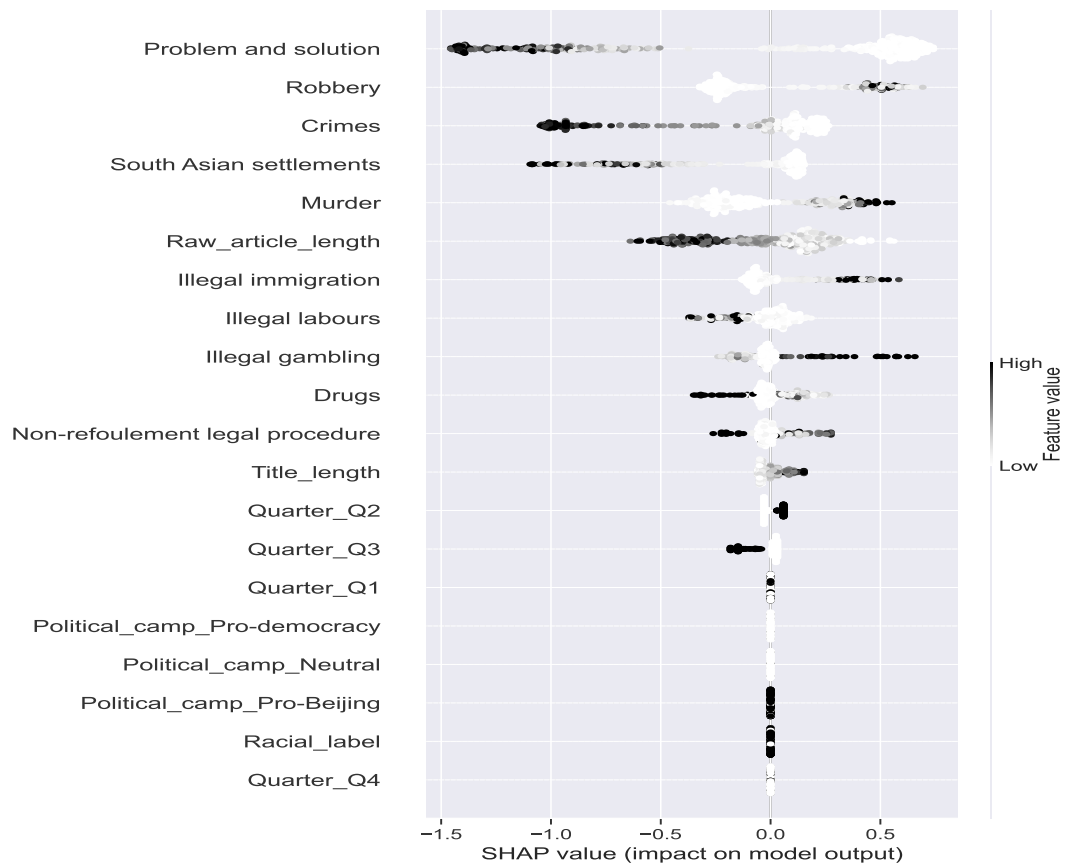


Figure 4.10: The SHAP values of the features in the prediction of whether an article has a neutral polarity

4. Results

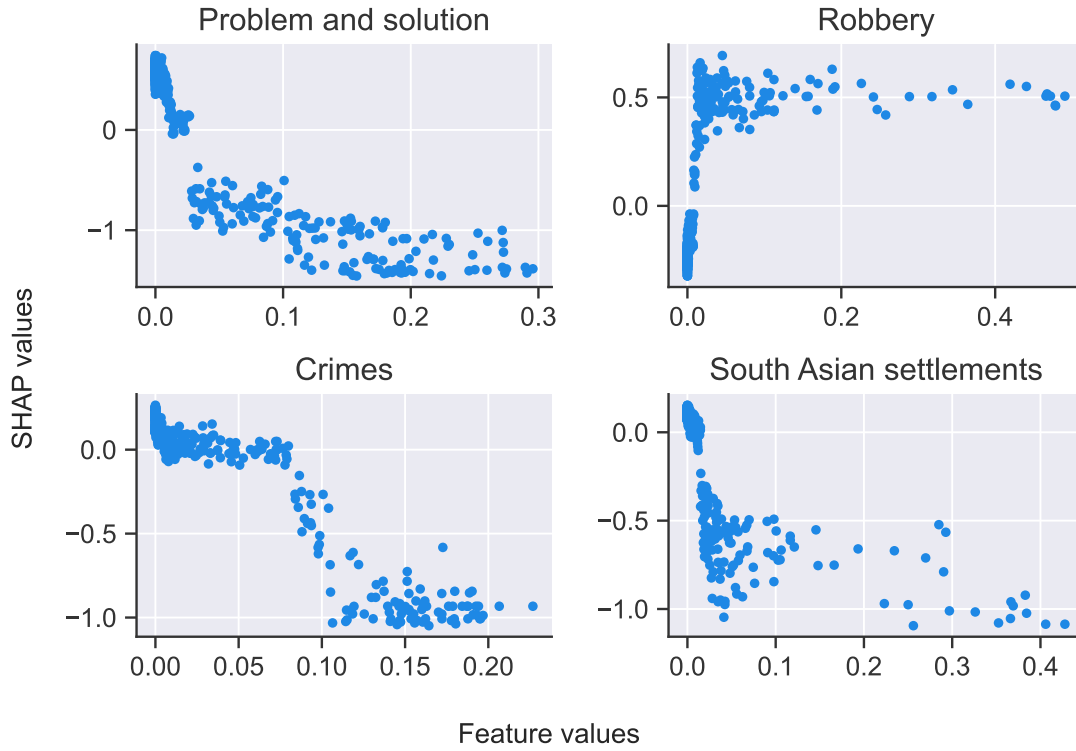


Figure 4.11: Dependence plot of the eight most important features for predicting neutral polarity of news articles

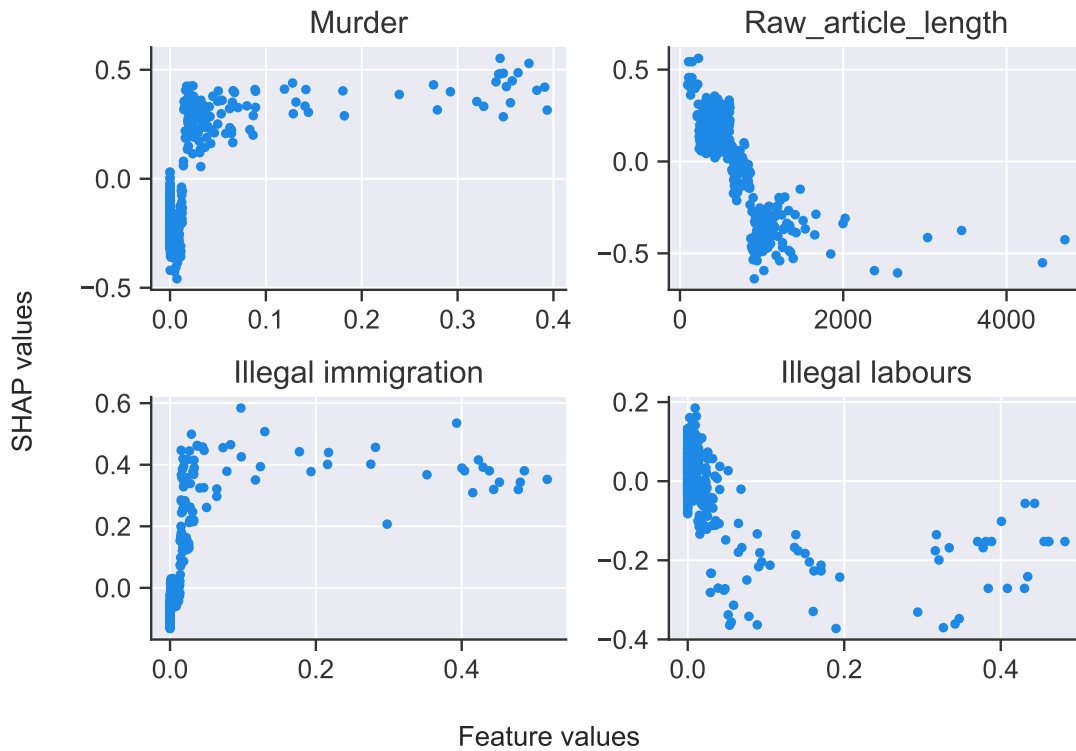


Figure 4.12: Dependence plot of the eight most important features for predicting neutral polarity of news articles

4. Results

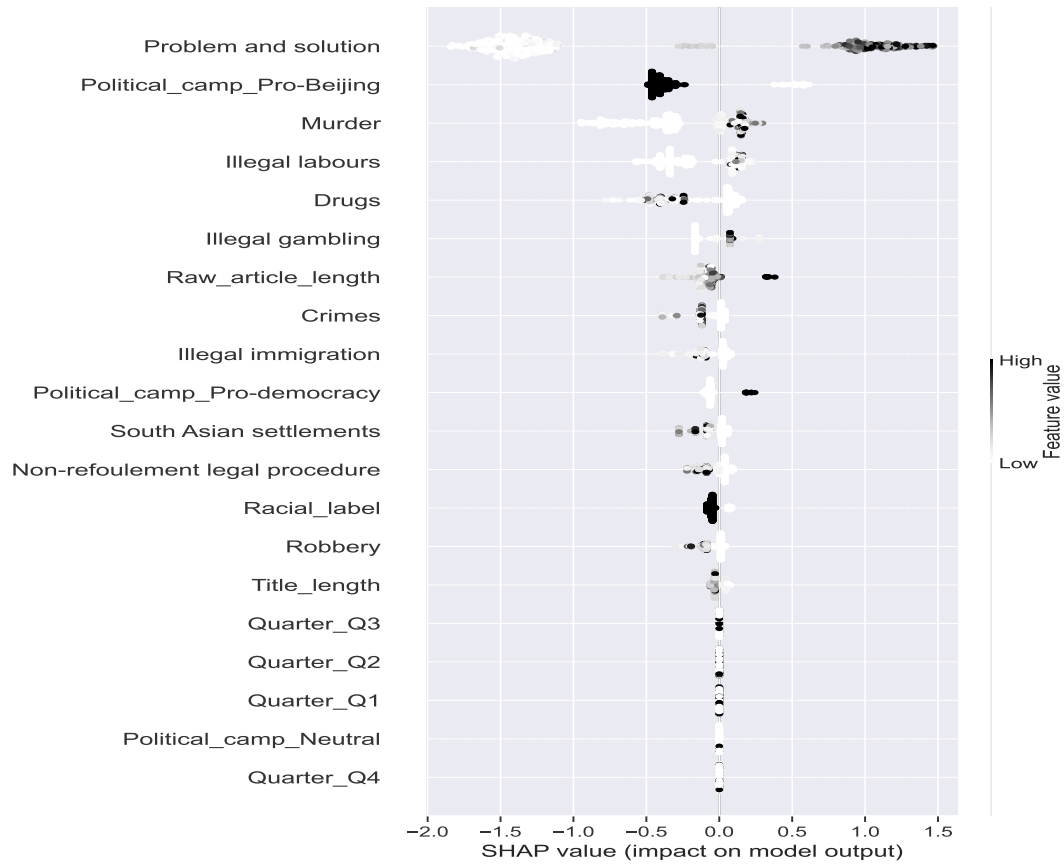


Figure 4.13: The SHAP values of the features in the prediction of whether an article has a positive polarity

newspapers in 2019. On the other hand, articles more related to Robbery, Illegal immigration and Murder were generally less loaded with sentiments by newspapers. As the length of the news articles (Raw_article_length) gets longer, it becomes more likely that the articles are loaded with conspicuous sentiments towards asylum seekers in Hong Kong.

4.3.4 SHAP values of predicting positive news articles

Figure 4.13 shows the SHAP values of the features in predicting whether news articles have a positive polarity towards non-refoulement claimants or not. This time, pro-Beijing affiliation (Political_camp_Pro-Beijing) is the second most crucial feature for predicting whether a news articles is likely to depict non-refoulement claimants positively. As expected in H_1 , a newspaper outlet is predicted to be less likely to publish positive articles about asylum seekers if it is associated with the pro-Beijing camp. By contrast, although with a smaller magnitude of SHAP values, pro-Democracy newspaper outlets are predicted to be more likely to have favourable

4. Results

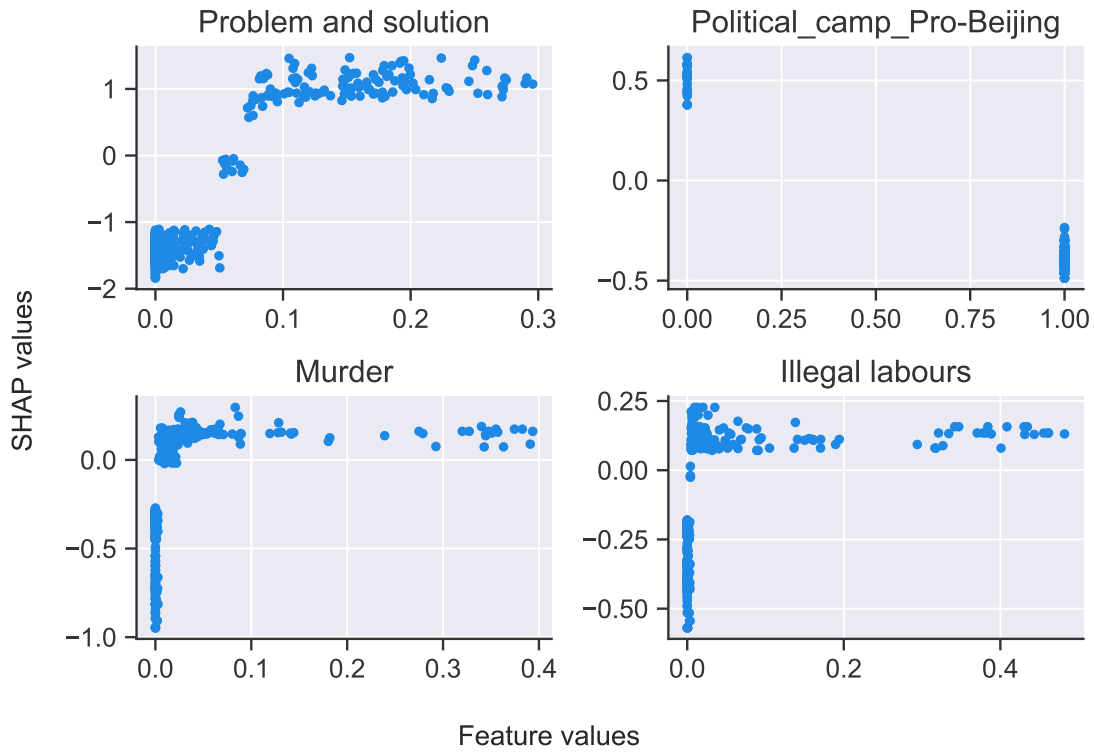


Figure 4.14: Dependence plot of the eight most important features for predicting positive polarity of news articles

reportage on non-refoulement claimants. Meanwhile, neutral newspaper outlets are neither more nor less likely to publish positive articles while reporting on asylum seekers in Hong Kong. In sum, pro-Beijing affiliation of newspaper media in Hong Kong is quite informative for predicting whether an article will portray non-refoulement claimants in positive light, and it is inferred from the model that pro-Beijing outlets are generally less likely to give positive coverage of this group of population.

Figures 4.14 and 4.15 again show the eight most important features on influencing the model to predict if a news article has positive polarity or not. Save for `problem and solution` (which will be explored later), quite a number of features (namely, `Murder`, `Illegal labours`, `Drugs`, `Illegal gambling` and `Crimes`) do not exhibit clear directions of the relationships between the feature and SHAP values, since the dependence plots of these features show rather “flat” trends of the correlation of these two values. Likewise, after omitting the few outliers in the upper-right quadrant on the dependence plot, `Raw_article_length` also do not show a very clear direction of how increasing the length of news articles may affect the SHAP values. The direction of

4. Results

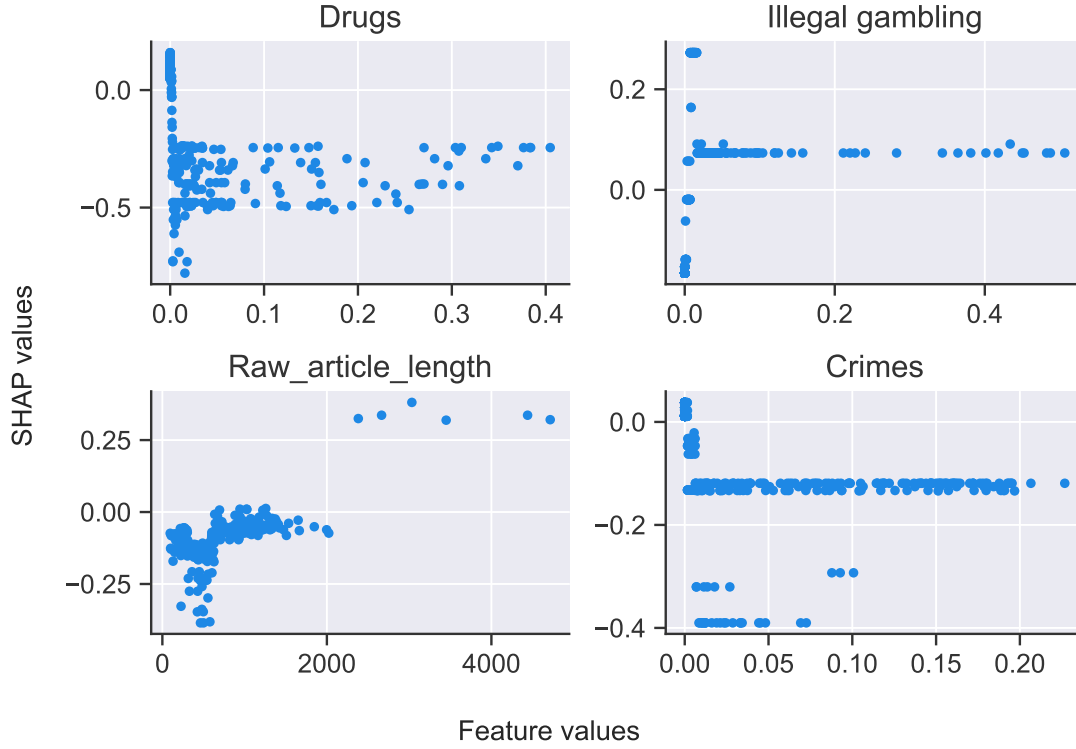


Figure 4.15: Dependence plot of the eight most important features for predicting positive polarity of news articles

the correlation between feature and SHAP values for the topic of `problem` and `solution` while predicting whether articles have positive polarity, however, merits further discussion because there appears to have an upward trend of this topic's SHAP values on the model's output as the feature value rises which is also the case when the model is predicting whether a newspaper article has a negative polarity. This will be discussed in the next section.

Before investigating the particularities surrounding the relationship between the feature and SHAP values of the `Problem` and `solution` topic, however, it is worth mentioning that due to the severe class imbalance problem for the dependent variable as shown in table 4.1, there are much fewer observations belonging to the positive polarity class in the dataset which will likely cause the variance of the predictions made on this class to be higher than those on the other two classes in the dependent variable. Accordingly, the findings about the associations between the feature and SHAP values of the `problem` and `solution` and `Political_camp_Pro-Beijing` features should be taken with a grain of salt because of the higher degree of uncertainty associated with the predictions on the positive class. Indeed, the lack of sufficient observations belonging

4. Results

to the positive class may also be the potential reason for the model's inability to discover the relationships between feature values and their impact on the model's output while predicting the positive class. That being said, the implication of the class imbalance problem goes beyond the modelling process because this can also be indicative of how Hong Kong's newspaper media cover the issue of non-refoulement claimants in general, and this will be elaborated further in the *Discussion* section.

4.3.5 Problem and solution: a topic with multiple angles by different camps?

The fact that the correlations between feature and SHAP values of the `Problem and solution` topic point to the same positive direction for the XGBoost model's prediction of both positive and negative polarities does not necessarily imply that the model is producing contradictory results. Instead, with many more possible way to be illustrated than other topics, it may be possible that the effect of `Problem and solution` in affecting the polarities of news articles is conditioned by other features (i.e. there exist interaction effects). This section will look deeper into this possibility with the help of dependence plots.

Based on the formulation of H_1 , figure 4.16 investigates whether there could be any interaction between the `Problem and solution` topic and the political affiliation of newspaper outlets (as expressed in whether they belong to the pro-Beijing camp or not). At first glance, both pro-Beijing and non-pro-Beijing newspaper outlets had news articles which emphasise on this topic (as shown by the black dots). Nevertheless, the within-group SHAP values of the news articles for each political camp are rather closely clustered together regardless of whether their news articles covered more or less on the `Problem and solution` topics. Ultimately, it was whether a newspaper outlet is affiliated with the pro-Beijing camp (on the x-axis) that clearly separated the SHAP values for the predictions of both negative and positive polarity of the news articles.

A potential reason for this observation is that each camp adopts different approaches and framing while writing on whether non-refoulement claimants in Hong Kong should be treated as a social problem and whether solutions to alleviate this "problem" are warranted. For instance, pro-Beijing newspaper outlets may follow suite the pro-Beijing politicians' hostility against non-refoulement claimants by constantly framing the latter as "fake refugees" (which is one of the topic words included in figure 4.5 for `Problem and solution`) to question their legitimacy of

4. Results

seeking refuge in Hong Kong in the first place while also alleging this group of population to multiple wrongdoings. By contrast, neutral newspaper outlets may be more likely to just report on news about asylum seekers in Hong Kong with little interpretation from the journalists, and pro-democracy outlets may even echo the stance of pro-democracy politicians in treating non-refoulement claimants more humanely while awaiting for the decision of their applications.

While qualitatively investigating how pro-Beijing and non-pro-Beijing newspaper outlets may use different framing to report on the topic of `Problem and solution` is out of the scope of this thesis, this can be a starting point for future research to prompt the dynamic of the contestation between pro-Beijing and pro-Democracy (and also the localist camp if possible) camps in attempting to persuade the public about how the issue of non-refoulement claimants in Hong Kong should be perceived. In particular, the anecdotal smearing of the localist camp's unwillingness to advocate more stringent measures against asylum seekers in Hong Kong by the pro-Beijing newspaper outlets (the Sun 2016) may signal that the issue of non-refoulement claimants may be exploited by political camps to criticise others of not protecting the interests of Hong Kong's society enough.

4.4 Discussion

Based on the results of SHAP values from the model, it can be concluded that H_1 is supported by the empirical evidence. In other words, a few years since the issue of non-refoulement claimants has become more visible in the public debate, pro-Beijing newspaper outlets were more likely to portray asylum seekers negatively than outlets with other political affiliations by 2019 after accounting for how the topics of the news articles might also affect the polarity of reportage. Moreover, certain topics (e.g. `Crimes`, `Non-refoulement legal procedure`) are more likely to portray asylum seekers negatively as they are more present in news articles, whereas the `Problem and solution` topic may be mentioned newspaper outlets from different political camps so differently that the higher presence of this topic in news articles is predicted by the model to increase the SHAP values of both positive and negative polarities.

There are also a few points worth mentioning based on the model's results. For starters, although the class imbalance problem of the `Sentiment` of the news articles on non-refoulement claimants in 2019 posed a few obstacles during the modelling process, this also suggests that

4. Results

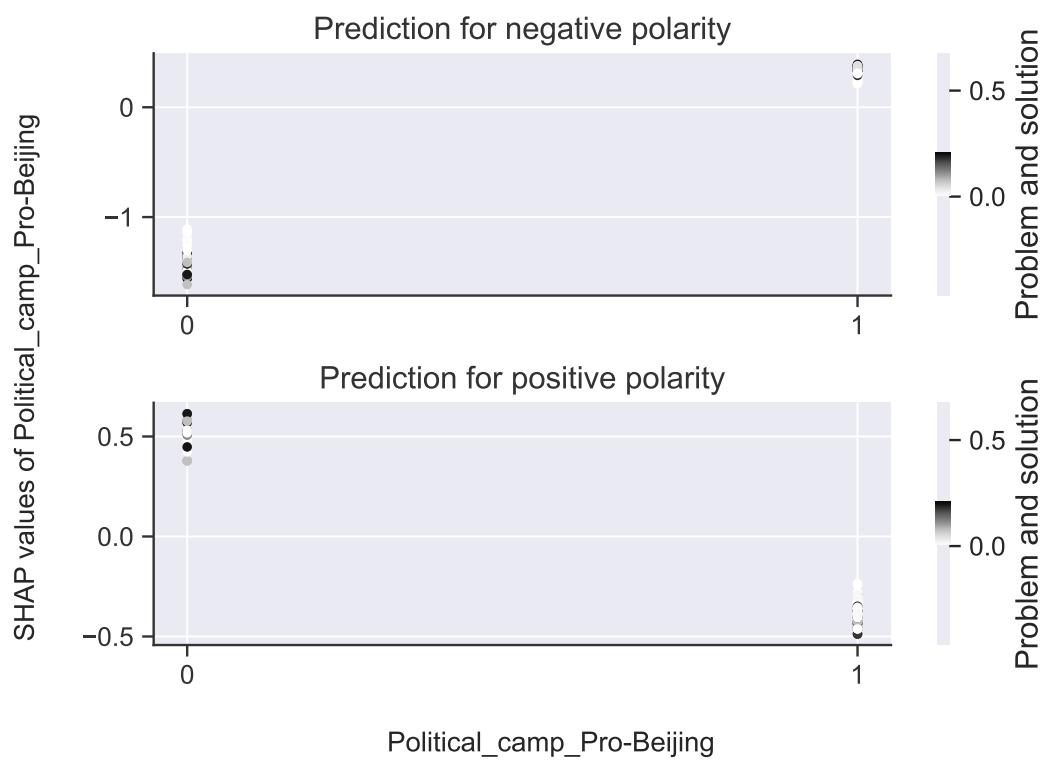


Figure 4.16: Interaction between pro-Beijing affiliation of newspaper outlets and the salience of the problem and solution topic for the XGBoost model's prediction of the negative (left) and positive (right) polarities

4. Results

printed newspapers in Hong Kong did not provide as much positive or even neutral coverage as it is the case for negative coverage. This phenomenon is likely contributed by the fact that *Oriental Daily News* was the sole major newspaper for publishing on the issue of asylum seekers in Hong Kong by 2019 while extensively using the term “fake refugee” among other derogatory descriptions to delegitimise the non-refoulement claimants. On the contrary, neutral and pro-democracy newspaper outlets did publish news about non-refoulement claimants at least without hostility and sometimes even included positive stories, but their combined volume of publication on this issue in 2019 was much smaller than that of pro-Beijing outlets altogether even after excluding those from *Oriental Daily News*. One can thus argue that the pro-Beijing camp was much more attentive than outlets with other political stances when it comes to reporting news about non-refoulement claimants. This is likely because the pro-Beijing camp would like to convince the public to agree that non-refoulement claimants, despite only constituting to a tiny proportion of Hong Kong’s total population of about 7.5 million according to figure 2.1, pose serious problems to the security and well-being of society, and thus pro-Beijing newspaper outlets frequently reported on news related to this group of population in negative light to reinforce the image that non-refoulement claimants are “troublemakers” which should not be welcomed in society.

Related to the publication volume by the pro-Beijing camp, a question about the validity of the results can be that the vast volume of publication by *Oriental Daily News* might have skewed the pro-Beijing media’s attitudes towards asylum seekers to the negative polarity. While it is certainly a fact that *Oriental Daily News* covered non-refoulement claimants disproportionately throughout 2019, dropping out news articles by *Oriental Daily News* may not be appropriate because this is essentially cutting out a huge part of the pro-Beijing camp’s discourse on non-refoulement claimants in the media industry from the analysis which will risk not reflecting the entirety of the camp’s media discourse. After all, H_1 is more interested in testing whether the sentiments of news articles towards non-refoulement claimants may differ *by political camp*, and *Oriental Daily News* is an integral part of the mainstream media network in Hong Kong co-opted by the Beijing Government (Lee 2018).

4. Results

Original:
定風波:港府勿再沽名釣譽

斬草不除根,春風吹又生。針對積重難返的南亞假難民問題.....
不管高官如何欲其加蓋之舌,大肆吹噓積壓的酷刑聲請個案有所減少,但市民面對的現實卻是,全港各區相繼被假難民大軍攻陷。
由深水埗通州街天橋底的南亞村,到整個油尖旺地區,再到新界的元朗屯門,概莫能外,香港再難尋得一片淨土。
日前有調查顯示,多達三分之一港人有移民意願,假難民肆虐導致市民難以「出入平安」,顯然是原因之一。
即使退一萬步說,港府花費大量資源應對假難民問題,先不說用於審核個案、打擊黑工等的人手,假難民的衣食住行、官司費用等,每年也要花費十億元以上公帑,一旦香港對假難民即捕即解,或設置禁閉式難民營,將相關社會成本用於幫助真正有需要的港人身上,能解決多少問題?

Translation:
Settlement of the storm: the Hong Kong Government should no longer sell its name and reputation

If you don't get rid of the grass, the wind will blow again in spring. About the recurring problem of South Asian fake refugees...
|
No matter how much the senior officials pontificate and boast about the decrease in the backlog of torture claims, the reality facing the public is that all districts in Hong Kong have been overrun by an army of fake refugees.

From South Asian Village at the bottom of the Tung Chau Street flyover in Sham Shui Po, to the entire Yau Tsim Mong area, to Tuen Mun, Yuen Long in the New Territories, nothing is left untouched in Hong Kong.

A few days ago, a survey showed that as many as one-third of Hong Kong people have the desire to emigrate, and it is obvious that the rampage of fake refugees makes it difficult for people to "enter and leave safely".

Even if we take a step back, the Hong Kong government spends a lot of resources to deal with the problem of fake refugees, not to mention the manpower used to examine cases and crack down on illegal workers, the clothing, food, housing and transportation of fake refugees, legal costs, etc., will cost more than a billion dollars of public money each year.

If Hong Kong arrested and discharged fake refugees while setting up detention camps, how many problems can be solved instead by spending the money on helping Hong Kong people in genuine need?

Figure 4.17: Excerpt from pro-Beijing Oriental Daily News

4.4.1 Snapshots of news articles

It will also be interesting to take some snapshots of how news articles from different political camps may report on non-refoulement claimants in Hong Kong. One way that the pro-Beijing newspapers portrayed non-refoulement claimants negatively in 2019 was accusing the latter of wasting public resources that could have been better used on Hong Kong citizens. For example, the abridged excerpt in figure 4.17 by Oriental Daily News describes non-refoulement claimants as “fake refugees” who are “overrunning” districts in Hong Kong as “armies” and wasting the Government’s public resources for processing their applications. Note how the first sentence of this article already connotes non-refoulement claimants with the South Asian racial group, and that their presence in many districts in Hong Kong has destroyed the peace of local citizens and motivated them to emigrate. The last two sentences also express discontent about the wasted resource on handling the “fake refugee” problem could have better used for improving the quality of life of citizens in need. In effect, such discourse by Oriental Daily News resembles how the majority ethnic-Chinese population is viewed as a family mentioned by Erni (2012), namely, the non-refoulement claimants from South Asia as outsiders both disrupt the way of life of the local citizens and prevent the Government from prioritising the interests of the local population.

Admittedly, some pro-Beijing outlets attempt to dissociate the so-called “fake refugees” from the ethnic minorities residing in Hong Kong, as the excerpt by *Ta kung Pao* in figure 4.18 blames the “bogus refugees” for tarnishing the reputation of the ethnic minorities who grew up in Hong

4. Results

Original:
假難民嚴重威脅社區治安
====
劉佩玉認為，目前的酷刑聲請問題亦會對香港的種族共融工作帶來極大負面影響，因為偽酷刑聲請來港的部分假難民，在破壞社區治安、環境的同時，亦使在港少數族裔人士的形象大打折扣，而這部分在香港土生土長的少數族裔人士，許多都幾代人以香港為家，認同香港社會的價值觀，是良好市民，一旦他們被標籤化，社區內難免產生芥蒂、誤會，不利於社會和諧。
====
Translation:
Fake Refugees Seriously Threaten Community Security
====
Liu Peiyu believes that the current problem of torture claims will also have a great negative impact on Hong Kong's racial integration efforts.

This is because some of the bogus refugees who come to Hong Kong on the basis of torture claims are damaging the law and order and environment of the community, and at the same time, the image of the ethnic minorities in Hong Kong is greatly degraded.

Many of these ethnic minorities, who were born and raised in Hong Kong, have called Hong Kong home for several generations, share the values of Hong Kong society, and are good citizens.
|
Once they are stigmatized, it is inevitable that there will be misunderstandings and misconceptions in the community, which is not conducive to social harmony.
====

Figure 4.18: Excerpt from pro-Beijing Ta Kung Pao

Original:
大狀：入境處對弱勢缺乏同情
====
【本報訊】曾收留斯諾登的香港三個難民家庭早前曾向港府申請酷刑聲請，全部被拒，其中獲加拿大批准難民身份的Vanessa在港生活近20年，她認為港府將申請庇護者當作香港的垃圾。在港難民沒有工作、沒有金錢、沒有希望和未來，有如生活在地獄之中。協助她的大律師田光響(Robert Tibbo，圖)向《蘋果》稱，加拿大的難民篩選制度遠較香港透明和公平，形容香港是「種族主義、歧視、排斥和迫害的世界」。港府審查制度不透明田光響向《蘋果》表示，Vanessa及其女兒在香港遭受歧視和邊緣化，她不被允許工作，亦沒有足夠的人道主義援助滿足其基本需求，其女兒更因膚色多次遭受種族歧視。
====
Translation:
Barristers: ImmI Lacks Compassion for the Disadvantaged
====
The three refugee families in Hong Kong who had taken in Snowden had earlier applied to the Hong Kong government for torture claims, all of which were rejected. One of them, Vanessa, who was granted refugee status by Canada, has been living in Hong Kong for nearly 20 years, and she thinks the Hong Kong government treats asylum seekers like Hong Kong's trash.

Refugees in Hong Kong have no jobs, no money, no hope or future, and are living in hell. Robert Tibbo (pictured), the barrister who assisted her, told Apple Daily that Canada's refugee screening system is far more transparent and fair than Hong Kong's, describing Hong Kong as "a world of racism, discrimination, exclusion and persecution."

The Hong Kong government's screening system is not transparent," Tibbo told The Apple, adding that Vanessa and her daughter are discriminated against and marginalized in Hong Kong, that she is not allowed to work, that there is not enough humanitarian assistance to meet her basic needs, and that her daughter has been repeatedly racially discriminated against because of her skin color.
====

Figure 4.19: Excerpt by pro-Democracy Apple Daily

Kong and are “good citizens”. However, it is unlikely that such separation can substantively reduce prejudices towards ethnic minorities in the sense that the coverage of Oriental Daily News is much more dominant in terms of volume even just within the pro-Beijing camp, and linkage between South Asian and “fake refugees” is frequently mentioned by this outlet which may be known and, in some cases accepted, by more readers.

On the other side, non-pro-Beijing newspaper outlets often cover non-refoulement claimants positively either by sharing their personal stories or advocating more sympathy towards this group of vulnerable population. Out of all the positive articles included in the dataset, the excerpt from Apple Daily in figure 4.19 is worth highlighting because it not only sympathetically reports the struggles faced by a former non-refoulement claimant in Hong Kong but also cites expert opinion which criticises the Hong Kong Government’s current asylum seeking regime of lacking transparency and fairness compared to the one in Canada. Such reportage aligns with the former pro-democracy legislators’ arguments that non-refoulement claimants should be treated more humane instead of potentially being indiscriminately detained under the accusation of being “fake refugees” as suggested by the pro-Beijing camp.

4. Results

4.4.2 Validity of automated sentiment analysis

Aside from the analytic findings, I should also mention the performance of using machine-learning-based methods for sentiment analysis to test H_1 . For starters, a satisfactory f1 score obtained by the tuned XGBoost model as shown in table 4.4 implies that the generated SHAP values are based on a performant model, thereby ensuring the validity of the interpretation in the *Results* section. Accordingly, this analysis demonstrates that natural language processing can be a powerful tool to scale research on larger datasets of text data in Chinese language where it would be very time-consuming to manually analyse every entry.

In addition to the performance of the model, NMF also proves to be a useful tool to summarise the topics of the articles in the dataset without resorting to pre-defined coding schemes. Specifically, the ten topics identified by the NMF model from this dataset are all semantically coherent, and only the **Problem** and **solution** topic has a wider scope of what content can be included than other topics. It is believed that this technique can prove quite useful for summarising the latent topics present in Chinese language corpora in the future, especially when there is no available coding schemes beforehand for classifying the topics of the documents, whether it be for reducing the dimensionality of the dataset or simply discovering the topics mentioned by the documents.

Chapter 5

Conclusion

To recap, I used XGBoost to build a multiclass classification model to investigate whether the sentiments (in the polarities of *negative*, *neutral* and *positive*) of the printed newspaper articles about non-refoulement claimants in Hong Kong may vary according to a newspaper's affiliated political camp. Because of the pro-Beijing camp's

It is found that after accounting for the content of the news articles (such as the dominant topics in a reportage) and other data of the articles (e.g. the length of the main text and whether racial labels were present), pro-Beijing newspaper outlets are both less likely to publish positive articles and more likely to write negative articles on asylum seekers in Hong Kong when compared to outlets with other political affiliations. While most of the topics

Besides the analysis results which aim to enrich the literature on media's portrayal of non-refoulement claimants in Hong Kong, to a large extent this study also

How can the case of Hong Kong on the

5.1 Limitations

Admittedly, this study suffers from a few limitations. First and foremost, because the articles included in the dataset did not cover online news media, the entries only included those published by printed newspapers in Hong Kong. Had I been able to scrap online news articles, it would have been possible to

5.2 Future research directions

In terms of methodology, future research on analysing the reportage

5.3 How might the instigation of the National Security Law affect the public discourse on asylum seekers in Hong Kong?

Just a year after the anti-extradition law protest had started and once again mobilised a huge section of Hong Kong's society against the authority, the HKSAR Government promulgated the National Security Law in July 2020 which aims to tighten the control over the political landscape of Hong Kong. As a result, non-pro-Beijing forces have witnessed their political influence being severely restricted, both in official institutions and the society. Even though the Legislative Council had been one of the few avenues where pro-democracy activists and politicians could run for office (despite lacking substantive authority to participate in policymaking by proposing bills), with the conclusion of the recent 2021 Legislative Council election after an overhaul of the electoral system which essentially permits candidacy only to the "patriots" loyal to the Beijing and HKSAR Governments (Lau and Yam 2021), the door for pro-democracy camp to advocate alternative political discourses has been shut. This means when pro-Beijing lawmakers propose non-binding bills to combat the so-called "fake refugee" problems as Holden Chow did back in 2016 (Cheng 2016a), such bills will very unlikely be opposed by others, and the voting results on the however unbinding bills may then be used by the Government to justify more stringent measures against non-refoulement claimants.

Even the presence of alternative media which strive to provide an alternative narrative of news contrary to the state-coopted mainstream media as Lee (2018) describes may not be the case anymore. In June 2021, the flagship pro-democracy newspaper *Apple Daily* was forced to shut down after the HKSAR Government had frozen the media outlet's asset under the pretext of violating the National Security Law (Tan 2021). Just days before the year 2021 ended, the popular pro-democracy online media *Stand News* had to shut down as well after the Hong Kong police raided its office and arrested senior staff (Ng and Pomfret 2021), followed by the closure of another online pro-democracy news outlets days after ("[Hong Kong](#)" 2022). It is likely that more alternative media may be pressured to defunct in the future as well due to the threat from the National Security Law, and this implies that alternative discourses will exist in the media industry of Hong Kong. Consequently, since the pro-Beijing camp media in general are more hostile to non-refoulement claimants, it is likely that Hong Kong's media landscape will be dominated by

5. Conclusion

more negative rhetoric towards asylum seekers in the city given that pro-democracy outlets are now finding increasing difficulties to operate without running into legal repercussion.

With two major channels of formulating and propagating political discourses being shut off, it can be expected that the pro-Beijing camp will face much less counter-discourse from the oppositions while attempting to frame socio-political issues, including non-refoulement claimants. Due to the lack of alternative sources of information to counter-balance

As the study by Boomgaarden and Vliegenthart (2009) shows, more pejorative coverage of refugees by the media may cause the public to be more averse towards this group of vulnerable population after some time, and the effect of negative news coverage may become even more salient when the levels of immigration and asylum seeking applications are perceived to be high.

In short, just as Hong Kong's political freedom is becoming narrower and narrower, hostility towards non-refoulement claimants may also increase in the foreseeable future after the Government has effectively eliminated any opportunity for the opposition to participate in policymaking however nominally. Whereas the media's rhetoric will be more and more dominated by pro-Beijing camp outlets as pro-democracy ones are closing

Appendix: The Echoes of the Code

The goal of this appendix is to echo the code you used in your thesis for a greater sense of transparency and replicability of your research. Note that `ref.labels` can be set to any label. Hence, you can filter the code you want replicated in the appendix by setting labels to the desired code chunks in the various chapters. See this excellent resource for more information.

This might be particularly useful when you perform model selection to output intermediary steps here instead of in the code to avoid cluttering your report.

Appendix: The Echoes of the Code redux

Add as many appendices as you like.

Below is the codes for setting up the hyper-parameter tuning grid for the XGBoost model. Note that the results from hyperopt may differ due to the stochastic nature of the search space defined for Bayesian Optimisation.

```
from hyperopt import fmin, hp, tpe, Trials
from sklearn.metrics import make_scorer, f1_score
import xgboost as xgb

# Setting up the hyper-parameter grid
xgb_space = {
    "n_estimators": hp.quniform("n_estimators", 10, 50, 5),
    "max_depth": hp.quniform("max_depth", 2, 8, 1),
    "learning_rate": hp.quniform("learning_rate", 0.01, 0.5, 0.01),
    "gamma": hp.quniform("gamma", 0.1, 10, 0.1),
    "min_child_weight": hp.quniform("min_child_weight", 1, 10, 1),
    "subsample": hp.quniform("subsample", 0.5, 0.9, 0.1),
    "colsample_bytree": hp.quniform("colsample_bytree", 0.5, 0.9, 0.1),
    "reg_lambda": hp.quniform("reg_lambda", 1, 100, 1)
}

# Defining the objective function
```

Appendix

```
def xgb_objective(params):  
    xgboost_clf = xgb.XGBClassifier(objective="multi:softmax",  
                                    eval_metric="mlogloss",  
                                    random_state=1,  
                                    use_label_encoder=False)  
  
    xgb_params = {  
        "n_estimators": int(params["n_estimators"]),  
        "max_depth": int(params["max_depth"]),  
        "learning_rate": params["learning_rate"],  
        "gamma": params["gamma"],  
        "min_child_weight": int(params["min_child_weight"]),  
        "subsample": params["subsample"],  
        "colsample_bytree": params["colsample_bytree"],  
        "reg_lambda": params["reg_lambda"]  
    }  
  
    loss = 1 - np.mean(cross_val_score(xgboost_clf.set_params(**xgb_params), X_train_final, y_train, cv=  
    return loss  
  
# Searching for optimal hyper-parameters  
xgb_trials = Trials()  
best_xgb_params = fmin(xgb_objective, xgb_space, algo=tpe.suggest, max_evals=1000, rstate=np.random.se  
  
# from sklearn.model_selection import cross_val_score, StratifiedKFold  
# from sklearn.metrics import classification_report  
  
# Utility function for evaluating the model's performance in cross validation and test set in terms of  
# def evaluate_model_f1(model, model_name: str, cv=five_fold_cv, X_train=X_train_final, X_test=X_test_  
#     y_pred = model.predict(X_test)  
#     cv_f1_score = np.mean(cross_val_score(model, X_train, y_train, cv=cv, scoring=make_scorer(f1_score  
#     test_f1_score = f1_score(y_test, y_pred, average="macro")
```

References

```
# return {"5-fold cv f1 score": cv_f1_score, "Test set f1 score": test_f1_score}

# Setting the appropriate data types of some hyper-parameters
best_xgb_params["n_estimators"] = int(best_xgb_params["n_estimators"])
best_xgb_params["max_depth"] = int(best_xgb_params["max_depth"])

# Fitting the model
xgb_tuned = xgb.XGBClassifier(objective="multi:softprob",
                              eval_metric="mlogloss",
                              random_state=1,
                              use_label_encoder=False).set_params(**best_xgb_params)

_ = xgb_tuned.fit(X_train_final,
                  y_train,
                  sample_weight=xgb_sample_weight,
                  eval_set=[(X_test_final, y_test)],
                  early_stopping_rounds=5,
                  verbose=0)

# Evaluating the tuned model
print(evaluate_model_f1(xgb_tuned, "Tuned xgboost"))
print(classification_report(y_test, xgb_tuned.predict(X_test_final)))

# To save the model locally. un-comment the below lines of code
import pickle
pickle.dump(xgb_tuned, open("xgb_clf_tuned_II.pkl", "wb"))
```

References

- Adintime (Apr. 2021). *The Most Widely-Read Magazine and Newspaper in Hong Kong*.
<https://adintime.hk/en/blog/the-most-widely-read-magazine-and-newspaper-in-hong-kong-n96>.
- Basic Law - Basic Law - Chapter I (EN) (Apr. 2021).
<https://www.basiclaw.gov.hk/en/basiclaw/chapter1.html>.
- BBC (July 2019). “Hong Kong: Protesters Storm and Deface Parliament on Handover Anniversary”. In: *BBC News*.
- BBC (Jan. 2021). *National Security Law: Hong Kong Rounds up 53 pro-Democracy Activists*.
<https://www.bbc.com/news/world-asia-china-55555299>.
- Boomgaarden, Hajo G. and Rens Vliegenthart (2009). “How News Content Influences Anti-Immigration Attitudes: Germany, 1993â€“2005”. In: *European Journal of Political Research* 48.4, pp. 516–542.
- Brownlee, Jason (July 2020). *Train-Test Split for Evaluating Machine Learning Algorithms*.
- Chen, Tianqi et al. (2015). “Xgboost: Extreme Gradient Boosting”. In: *R package version 0.4-2* 1.4, pp. 1–4.
- Cheng, Kris (Dec. 2016a). *Legislature Rejects Lawmaker’s Non-Binding Motion to Combat ‘bogus’ Refugees*. <https://hongkongfp.com/2016/12/01/legislature-rejects-lawmakers-non-binding-motion-combat-bogus-refugees/>.
- Cheng, Kris (Jan. 2016b). *NGOs Hit Back as CY Says Gov’t May Quit UN Torture Convention to Block ‘fake’ Refugees*. <https://hongkongfp.com/2016/01/14/ngos-hit-back-as-cy-says-govt-may-quit-un-torture-convention-to-block-fake-refugees/>.
- Cheung, Karen and Tom Grundy (Apr. 2016). *DAB Party Sets up in Tsim Sha Tsui Calling for ‘fake’ Refugees to Be Put in Camp and Deported*. <https://hongkongfp.com/2016/04/11/dab-party-sets-up-in-tsim-sha-tsui-calling-for-fake-refugees-to-be-put-in-camp-and-deported/>.

References

- Chun, Allen (1996). "Fuck Chineseness: On the Ambiguities of Ethnicity as Culture as Identity". In: *boundary 2* 23.2, p. 111. ISSN: 01903659. DOI: 10.2307/303809.
- Cooper, Samantha et al. (2017). "Media Coverage of Refugees and Asylum Seekers in Regional Australia: A Critical Discourse Analysis". In: *Media International Australia* 162.1, pp. 78–89.
- CUHK (2020). *Public Evaluation on Media Credibility*.
<https://ccpos.com.cuhk.edu.hk/public-evaluation-on-media-credibility/>.
- Department, Immigration (2021a). *Enforcement / Statistics on Non-refoulement Claim*.
<https://www.immd.gov.hk/eng/facts/enforcement.html>.
- Department, Immigration (2021b). *Making a Claim for Non-refoulement Protection in Hong Kong*.
https://www.immd.gov.hk/eng/useful_information/non-refoulement-making-claim.html.
- elgarteo (2021). *HKPolDict: Hong Kong Politics Dictionary for Text Mining*.
<https://github.com/elgarteo/HKPolDict>.
- Erni, John Nguyet (2012). "Who Needs Strangers? Un-imagining Hong Kong Chineseness". In: *Chinese Journal of Communication* 5.1, pp. 78–87.
- Fong, Brian (2017a). "In-between Liberal Authoritarianism and Electoral Authoritarianism: Hong Kong's Democratization under Chinese Sovereignty, 1997–2016". In: *Democratization* 24.4, pp. 724–750. ISSN: 1351-0347, 1743-890X. DOI: 10.1080/13510347.2016.1232249.
- Fong, Brian CH (2017b). "One Country, Two Nationalisms: Center-periphery Relations between Mainland China and Hong Kong, 1997–2016". In: *Modern China* 43.5, pp. 523–556.
- Full Text of Constitution of Communist Party of China - Xinhua* (Nov. 2017).
http://www.xinhuanet.com/english/special/2017-11/03/c_136725945.htm.
- Greenfeld, Liah and Jonathan Eastwood (2007). "Nationalism and National Identity: Definitional Issues". In: *The Oxford Handbook of Comparative Politics*. Ed. by Carles Boix and Susan Stokes. 1st. Oxford University Press, pp. 256–273.
- Heath, Anthony F. and James R. Tilley (2005). "British National Identity and Attitudes towards Immigration". In: *International Journal on Multicultural Societies* 7.2, pp. 119–132.
- Hitchens, Dan (July 2021). "Opinion | The Christian Heart of Hong Kong Activism". In: *Wall Street Journal*. ISSN: 0099-9660.

References

Hjerm, Mikael (1998). "National Identity: A Comparison of Sweden, Germany and Australia".

In: *Journal of Ethnic and Migration Studies* 24.3, pp. 451–469.

Home_Affairs_Department (Jan. 2018). *The Demographics: Ethnic Groups*.

https://www.had.gov.hk/rru/english/info/info_dem.html.

"Hong Kong" (Jan. 2022). "Hong Kong: Pro-democracy Website Citizen News Is Latest to Close". In: *BBC News*.

Hui4yi4 Guo4cheng2 Zheng4shi4 Ji4lu4 (Dec. 2016).

<https://webcache.googleusercontent.com/search?q=cache:zyvAUu3wI5oJ:https://www.legco.gov.hk/yr16-17/chinese/counmtg/hansard/cm20161201-translate-c.pdf+&cd=8&hl=zh-TW&ct=clnk&gl=ch>.

Junyi, Sun (Dec. 2021). *Jieba*.

Kaeding, Malte Philipp (2017). "The Rise of" Localism" in Hong Kong". In: *Journal of Democracy* 28.1, pp. 157–171.

Kwan, Justin P. (2016). "The Rise of Civic Nationalism: Shifting Identities in Hong Kong and Taiwan". In: *Contemporary Chinese Political Economy and Strategic Relations: An International Journal* 2.2, pp. 941–973.

Kwong, Ying-ho (2016). "The Growth of "Localism" in Hong Kong. A New Path for the Democracy Movement?" In: *China Perspectives* 2016.2016/3, pp. 63–68.

Lau, Jessie and Shui-yin Sharon Yam (Dec. 2021). *'Patriots Only': Hong Kong's New Election System in Action*.

<https://thediplomat.com/2021/11/patriots-only-hong-kongs-new-election-system-in-action/>.

Lau, Tracy (2013). "State Formation and Education in Hong Kong: Pro-Beijing Schools and National Education". In: *Asian Survey* 53.4, pp. 728–753.

Laver, Michael, Kenneth Benoit, and John Garry (2003). "Extracting Policy Positions from Political Texts Using Words as Data". In: *American political science review* 97.2, pp. 311–331.

Law, Kam-Yee and Kim-Ming Lee (Mar. 2012). "The Myth of Multiculturalism in 'Asia's World City': Incomprehensive Policies for Ethnic Minorities in Hong Kong". In: *Journal of Asian Public Policy* 5.1, pp. 117–134. ISSN: 1751-6234. DOI: 10.1080/17516234.2012.662353.

References

- Lecours, André (2000). "Ethnic and Civic Nationalism: Towards a New Dimension". In: *Space and Polity* 4.2, pp. 153–166.
- Lee, Francis LF (2018). "Changing Political Economy of the Hong Kong Media". In: *China Perspectives* 2018.2018/3, pp. 9–18.
- Ler, Wei Han (Mar. 2021). *High Number of Unique Values and Tree Based Models*.
<https://towardsdatascience.com/high-number-of-unique-values-and-tree-based-models-c30167b63edd>.
- Lewin-Epstein, Noah and Asaf Levanon (2005). "National Identity and Xenophobia in an Ethnically Divided Society". In: *International Journal on Multicultural Societies* 7.2, pp. 90–118.
- Lin, Cong and Liz Jackson (2021). "Make China Great Again: The Blood-Based View of Chineseness in Hong Kong". In: *Educational Philosophy and Theory* 53.9, pp. 907–919.
- Loh, Christine (2010). *Underground Front: The Chinese Communist Party in Hong Kong*. Hong Kong University Press.
- Lundberg, Scott and Su-In Lee (Nov. 2017). "A Unified Approach to Interpreting Model Predictions". In: *arXiv:1705.07874 [cs, stat]*. arXiv: 1705.07874 [cs, stat].
- Ma, Ngok (2012). "Political Parties and Elections". In: *Contemporary Hong Kong Government and Politics*. Ed. by Wai-man Lam, Percy Luen-tim Lui, and Wilson Wong. Second. Hong Kong University Press, pp. 159–176.
- Manifesto* (2021). https://www.dphk.org/index.php?route=eng/about_us/detail&id=1.
- Marr, Bernard (May 2018). *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*. <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/>.
- Mathews, Gordon (2018). "Asylum Seekers as Symbols of Hong Kong's Non-Chineseness". In: *China Perspectives* 2018.2018/3, pp. 51–58.
- McKay, Fiona H., Samantha L. Thomas, and R. Warwick Blood (2011). "'Any One of These Boat People Could Be a Terrorist for All We Know!' Media Representations and Public Perceptions of Boat People Arrivals in Australia". In: *Journalism* 12.5, pp. 607–626.
- Medrano, Juan Diez (2005). "Nation, Citizenship and Immigration in Contemporary Spain". In: *International Journal on Multicultural Societies* 7.2, pp. 133–156.

References

- Medrano, Juan Diez and Matthias Koenig (2005). "Nationalism, Citizenship and Immigration in Social Science Research". In: *International Journal on Multicultural Societies* 7.2, pp. 82–89.
- Ng, Edmond and James Pomfret (Dec. 2021). "Hong Kong Pro-Democracy Stand News Closes after Police Raids Condemned by U.N., Germany". In: *Reuters*.
- Ng, Isabella (2020). "Criminalizing the Innocents: Social Exclusion of Asylum-Seekers and Refugees in Hong Kong". In: *Journal of Asian Public Policy* 13.3, pp. 319–332.
- Ng, Isabella, Sharice Fung-Yee Choi, and Alex Lih-Shing Chan (2021). "Resistance to 'Framing'? The Portrayal of Asylum Seekers and Refugees in Hong Kong's Online Media". In: *Journalism Practice*, pp. 1–17.
- Ng, Isabella, Sharice Fungyee Choi, and Alex Lihshing Chan (May 2019). "Framing the Issue of Asylum Seekers and Refugees for Tougher Refugee Policy—a Study of the Media's Portrayal in Post-colonial Hong Kong". In: *Journal of International Migration and Integration* 20.2, pp. 593–617. ISSN: 1874-6365. DOI: 10.1007/s12134-018-0624-7.
- Ngo, Jennifer (Apr. 2016). *Political Dark Arts Live on in Hong Kong with Bogus Banners Designed to Shame Lawmakers*. <https://www.scmp.com/news/hong-kong/politics/article/1933095/political-dark-arts-live-hong-kong-bogus-banners-designed>.
- O'Doherty, Kieran and Amanda Lecouteur (2007). "'Asylum Seekers', 'Boat People' and 'Illegal Immigrants': Social Categorisation in the Media". In: *Australian journal of psychology* 59.1, pp. 1–12.
- OCTOBER ROUNDUP AT THE REFUGEE UNION (Nov. 2018). <http://refugeeunion.org/9739/october-roundup-refugee-union-office/>.
- Pedregosa, Fabian et al. (2011). "Scikit-Learn: Machine Learning in Python". In: *the Journal of machine Learning research* 12, pp. 2825–2830.
- Po, Kung Kao (Oct. 2019). *chen2 shu1ji1 yu3 xin4tu2 xian4 mi2sa5 ji4nian4 yu3san3 yun4dong4 wu3 zhou1nian2*. <https://kkp.org.hk/node/19587>.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Refugee Community Supports Lawmaker Fernando Cheung Chiu-hung (Mar. 2015). <http://www.vfnw.org/7401/vision-first-supports-lawmaker-fernando-cheung-chiu-hung/>.

References

- Refugee Population by Country or Territory of Asylum - Hong Kong SAR, China / Data* (2021).
<https://data.worldbank.org/indicator/SM.POP.REFG?locations=HK>.
- Refugees, Asylum-seekers and Migrants* (2022).
<https://www.amnesty.org/en/what-we-do/refugees-asylum-seekers-and-migrants/>. Amnesty International.
- Shap.TreeExplainer SHAP Latest Documentation* (n.d.).
https://shap-lrjball.readthedocs.io/en/docs_update/generated/shap.TreeExplainer.html.
- So, Alvin Y. (2015). “The Making of Hong Kong Nationalism”. In: *Asian Nationalisms Reconsidered*. Ed. by Jeff Kingston. 1st. Routledge, pp. 151–162.
- Soltoff, Benjamin (Sept. 2021). *Topic Modeling*.
<https://cfss.uchicago.edu/notes/topic-modeling/>.
- Stevens, Keith et al. (2012). “Exploring Topic Coherence over Many Models and Many Topics”. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 952–961.
- Tan, Yvette (June 2021). “Apple Daily: The Hong Kong Newspaper That Pushed the Boundary”. In: *BBC News*.
- the Sun (Mar. 2016). *jiu4shi4lun4shi4: ben3tu3pai4 shi1sheng1*.
http://the-sun.on.cc/cnt/news/20160312/00418_009.html.
- Theme of the 32nd Anni. of June 4 “For Freedom, a Common Fate, a Shared Struggle”* (2021).
<https://web.archive.org/web/20210501051630/https://hka8964.wordpress.com/6432-2/>.
- Unified Screening Mechanism Procedures* (2021).
<https://www.unhcr.org/hk/en/2675-usm-procedures-2.html>.
- van Atteveldt, Wouter, Mariken A. C. G. van der Velden, and Mark Boukes (Apr. 2021). “The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms”. In: *Communication Methods and Measures* 15.2, pp. 121–140. ISSN: 1931-2458. DOI: 10.1080/19312458.2020.1869198.
- Varshney, Ashutosh (2007). “Ethnicity and Ethnic Conflict”. In: *The Oxford Handbook of Comparative Politics*. Ed. by Carles Boix and Susan Stokes. 1st. Oxford University Press, pp. 274–294.
- Vecchio, Francesco (2014). *Asylum Seeking and the Global City*. Routledge.

References

Voices for Protection Go to LegCo (July 2014).

Xie, Yihui (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*.

ISBN 978-1138700109. Boca Raton, Florida: Chapman and Hall/CRC. URL:

<https://bookdown.org/yihui/bookdown>.

Xie, Yihui (2021). *bookdown: Authoring Books and Technical Documents with R Markdown*. R

package version 0.24. URL: <https://CRAN.R-project.org/package=bookdown>.

Yuen, Chantal (Nov. 2016). *Pro-Beijing Lawmaker Submits Motion to Stop 'Bogus Refugees'*.

<https://hongkongfp.com/2016/11/18/pro-establishment-lawmaker-submits-motion-to-stop-bogus-refugees/>.