When central-peripheral relations intersect with the asylum-seeking regime: how media outlets from different political camps report on asylum seekers 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. 12345

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

Geneva
2022

INSTITUT DE HAUTES ETUDES INTERNATIONALES ET DU DEVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

RESUME / ABSTRACT

Non-refoulement claimants (or asylum seekers) has become a salient issue 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 ideologies saliently, how might the depiction of non-refoulement claimants vary according to the political camp that media outlets were 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 different political stances. With the promulgation of the National Security Law in 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 causing it to become more hostile against those who seek refuge in Hong Kong due to facing threats of tortures in their home countries.

For my loved ones

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.

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.

On a personal level, my beloved parents have offered their unlimited support to me throughout my study at the Institute and my life in general. 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.

<div align="right">

Chin Man KWAN

Graduate Institute, Geneva

10 January 2022

</div>

Contents

List of Figures

List of Tables

List of Abbreviations

ASPDMC   the Alliance for the Support of the Patriotic and Democratic Movement

NATO   North Atlantic Treaty Organization

UN   United Nations

UNICEF   United Nations Children's Fund

WHO   World Health Organization

WTO   World Trade Organization

Preface

Perhaps one of the most salient challenges facing the world in the 21st Century will be the massive movement of asylum seekers and refugees across national borders who are forced to flee from their home countries due to threats from persecutions and human rights violations (International 2022), and the 2015 Refugee Crisis in Europe is undoubtedly one of the most high-profile incidents on this issue. Political scientists have continuously inquired the relationship between identity politics and the attitudes towards immigrants (of which asylum seekers are a subgroup) of the public and policymakers, with a considerable amount of work focusing on the European context (e.g. Börzel and Risse (2018); Noury and Roland (2020)). There certainly exists various entry points for one to investigate the politics related to asylum seeking, but the mass media's rhetoric on asylum seekers is often under academic scrutiny because this is regarded as one of the most prominent ways for potentially influencing public perceptions of asylum seekers as per the literature, given mass media's role as one of the most accessible and popular sources of information for the public to know about affairs of the world (Cooper et al. 2017; McKay et al. 2011; Ng, Choi, et al. 2019) .

How do these two phenomena inspired me to write about how the media portray asylum seekers (or more formally, non-refoulement claimants) residing in Hong Kong? Starting with the issue of asylum seeker *per se*, Hong Kong has been a destination for refuge for the second half of the 20th Century (first for those fleeing from mainland China during the Maoist period and then Vietnamese facing threats of persecutions after the Vietnam War).

The media landscape in Hong Kong

Even though Hong Kong is never a signatory of the 1951 Refugee Convention and thus is not obligated to host refugees within its territory, the Government still has to provide refuge to those who flee from their countries of origin due to

My inspiration for writing on this topic comes from several sources. For starters, .

Furthermore, I would like to take the opportunity of researching on this important topic to test whether machine learning

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 asylum seekers based on numerous perspectives and levels of analysis, it was quite easy to often get lost. 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 the political division in Hong Kong is arguably more defined based on how the identity of Hong Kong citizens should be defined, even though it is equally valid to classify the political parties according to the more conventional left-right dimensions (Ma 2012).

The data collection and preprocessing part also took quite a considerable amount of time. Apart from only filtering out relevant news articles about non-refoulement claimants published by newspapers in 2019, the messy nature of text data also required a lot of steps to transform them into a format suitable for being model inputs. The fact that the data set did not have

As much as there are numerous performant machine learning models available nowadays, the result will not be very meaningful if the data quality is not optimal.

Chapter 1

Introduction

Asylum seekers, refugees and these two groups of vulnerable population's movements across national borders have become some of the most salient issues in the 21st Century, especially after the outbreak of the 2015 Refugee Crisis in Europe. A lot of academic researches have subsequently investigated how politics in the .

Since the transfer of sovereignty from Britain to China in 1997, Hong Kong has witnessed increasing tension with its new sovereign owner. Until July 2020 when the HKSAR[1] 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 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.

This thesis thus aims to investigate ***how may the framing of asylum seekers in Hong Kong vary by news media outlets associated with different political camps, given each of them has considerably different definition of the city's national identity.*** The academic contribution of this thesis .

Indeed, the influence of the media on

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

---

[1]For 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.

₂₂₂ definitions of asylum seekers and refugees[2]. Nevertheless, since the above three terms are generally

₂₂₃ used to refer to the same group of population in Hong Kong who flee from their own countries

₂₂₄ to escape from persecution, treating these terms as synonyms in this context should better cover

₂₂₅ how this group of population is perceived by media outlets in different political camps. In fact,

₂₂₆ scholars researching on this issue often use the terms I listed without strictly differentiating their

₂₂₇ meanings (e.g. Vecchio (2014), Ng, Choi, et al. (2019)).

₂₂₈ This thesis will proceed as follow. In section 2, I will review the literature on the theories

₂₂₉ about how the perception of national identity might affect attitudes towards immigration, how

₂₃₀ each major political camp in Hong Kong generally perceives the national identity of Hong Kong,

₂₃₁ and how each camp perceives asylum seekers. Section 3 provides an overview of the data collection

₂₃₂ process and methodology used to answer the research questions. Section 4 presents the result

₂₃₃ of the analysis, and section 5 concludes the findings as well as mention the implications of this

₂₃₄ thesis.

---

[2]The code for tuning the model can be found in the appendix. Hyper-parameters of the tuned 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.

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

5

²⁵⁷ ultimately control political legitimacy and authority. These features together contribute to two

²⁵⁸ fundamental implications. Socially, the fact that nationals are considered as being equal and that

²⁵⁹ status is achieved instead of ascribed promotes social mobility. Politically, nationalism has made

²⁶⁰ modern (nation-)states "impersonal" (p.261) because whoever holds the government office does

²⁶¹ not affect whatsoever the nation-state's existence and legitimacy which ultimately rest in popular

²⁶² sovereignty (Greenfeld and Eastwood 2007). But for cultures that had been existing long before

²⁶³ nationalism's entry into these societies, the constructions of novel national identities may often

²⁶⁴ rebrand pre-existing cultures despite the latter's internal heterogeneity, of which Chineseness is

²⁶⁵ an example which I will look into later.

²⁶⁶ Lastly, Greenfeld and Eastwood (2007) propose a typology of nationalism with two criteria.

²⁶⁷ The first one is how the nation-state relates itself to its members, from the individualist end

²⁶⁸ of treating a nation as merely a massive group of individuals to the collectivist end of which

²⁶⁹ individuals are subservient to the nation. The other criterion is the well-known "ethnic" versus

²⁷⁰ "civic" dimension which is extensively studied, including its relation with the perception towards

²⁷¹ immigrants which I will discuss later. In general, ethnic nationalism defines a nation's members

²⁷² based on ascriptive and rather objective standards such as descent, language and customs which

²⁷³ are largely decided at birth, whereas civic nationalism focuses less on the members' cultural

²⁷⁴ backgrounds than their voluntary allegiance to a political community and its legal institutions

²⁷⁵ within a political territory (Greenfeld and Eastwood 2007; Lecours 2000; Lewin-Epstein and

²⁷⁶ Levanon 2005; Heath and Tilley 2005).

²⁷⁷ A note on the ethnic-civic dimension is that these two aspects are not mutually exclusive since

²⁷⁸ Medrano and Koenig (2005) highlight that nationalism in one country can often incorporate both

²⁷⁹ civic and ethnic features, which is quite frequently observed from empirical. Accordingly, scholars

²⁸⁰ often create other dimensions of describing nationalism theoretically. For example, Hjerm (1998)

²⁸¹ proposes to additionally gauge whether identifications with ethnic and civic features of nationalism

²⁸² are high ("multiple national identity") or low ("pluralist") at the same time (p. 453). Meanwhile,

²⁸³ some scholars prefer more fine-grained categories of citizenship criteria. Medrano (2005) breaks

²⁸⁴ the ethnic and civic dimensions further into five categories while describing the importance of

²⁸⁵ criteria for citizenship perceived by the Spanish population, namely (p.137),

²⁸⁶ • the civic dimension being broken down into:

- – 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)

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.

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

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 Kuomingtang (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 maindland. 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

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 How do different political camps in Hong Kong perceive national identity?

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 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 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 poltician 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; 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

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 note that the pro-democracy camp in general still identify 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.

⁵⁵⁰ The geographical scope of how wide Hong Kong's political movement to democratisation should

⁵⁵¹ concern, then, is partly what makes the localist camp different from the pro-democracy camp

⁵⁵² even though both favour a democratic regime in Hong Kong.

The localist camp

⁵⁵⁴ Broadly speaking, the rise of localism as a third prominent political ideology in the 2010s was due

⁵⁵⁵ to both the increasing integration with mainland China both politically and socio-economically

⁵⁵⁶ as well as the perceived incompetence of the pro-Democracy camp to push forward substan-

⁵⁵⁷ tive democratisation in Hong Kong. Although the term "localist camp" in reality represents an

⁵⁵⁸ aggregation of groups which vary considerably in choosing violent or non-violent means of re-

⁵⁵⁹ sisting against the Beijing Government, whether Hong Kong should abandon OCTS and pursue

⁵⁶⁰ independence, and are much less coordinated in their actions compared to the pro-Beijing and

⁵⁶¹ pro-democracy camps, organisations belonging to this camp hold strong averse sentiment against

⁵⁶² China **both** in terms of the PRC regime and mainland Chinese who are deemed to be depriving

⁵⁶³ the locals of already scarce public resources or disrupting the way of life of ordinary Hong Kong

⁵⁶⁴ citizens, and they (Kwong 2016) . The first substantive discourse to materialise localism as a

⁵⁶⁵ political ideology was "Hong Kong as a city-state" published in 2011 by scholar *Chin Wan-kan*

⁵⁶⁶ *which*, among many other ideas, advocates separating the integration between Hong Kong and

⁵⁶⁷ mainland China as well as prioritising the needs of Hongkongers (So 2015) . The following para-

⁵⁶⁸ graphs will outline how the rise of localism can be treated as a backlash against both long-existing

⁵⁶⁹ political camps in Hong Kong while also exploring how localists perceive Hong Kong's national

⁵⁷⁰ identity.

⁵⁷¹ Starting with the so-called "China factor", the localist camp obviously rejects the pro-Beijing

⁵⁷² camp's (and the Chinese Government) claim that Hong Kong's identity should be treated as one

⁵⁷³ with that of Chinese. Fong (2017b) illustrates how the China factor has facilitated the rise of

⁵⁷⁴ localism under the framework of "state-building" versus "peripheral" nationalism (p.2), where

⁵⁷⁵ Hong Kong is attempting to consolidate its unique identity against Beijing's incorporation of

⁵⁷⁶ the city into the hegemonic national culture. Specifically, after 500,000 Hongkongers had rallied

⁵⁷⁷ against the Article 23 legislation in 2003, Beijing then started to increase its interventions in

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,

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

	Pro-Beijing camp	Pro-Democracy camp	Localist camp
The nation that Hong Kong belongs to	China	China	Hong Kong
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

2.2.2 Asylum seekers in Hong Kong

Hong Kong's 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 re-iterates 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" (*Making a Claim for Non-refoulement Protection in Hong Kong 2021*) , 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
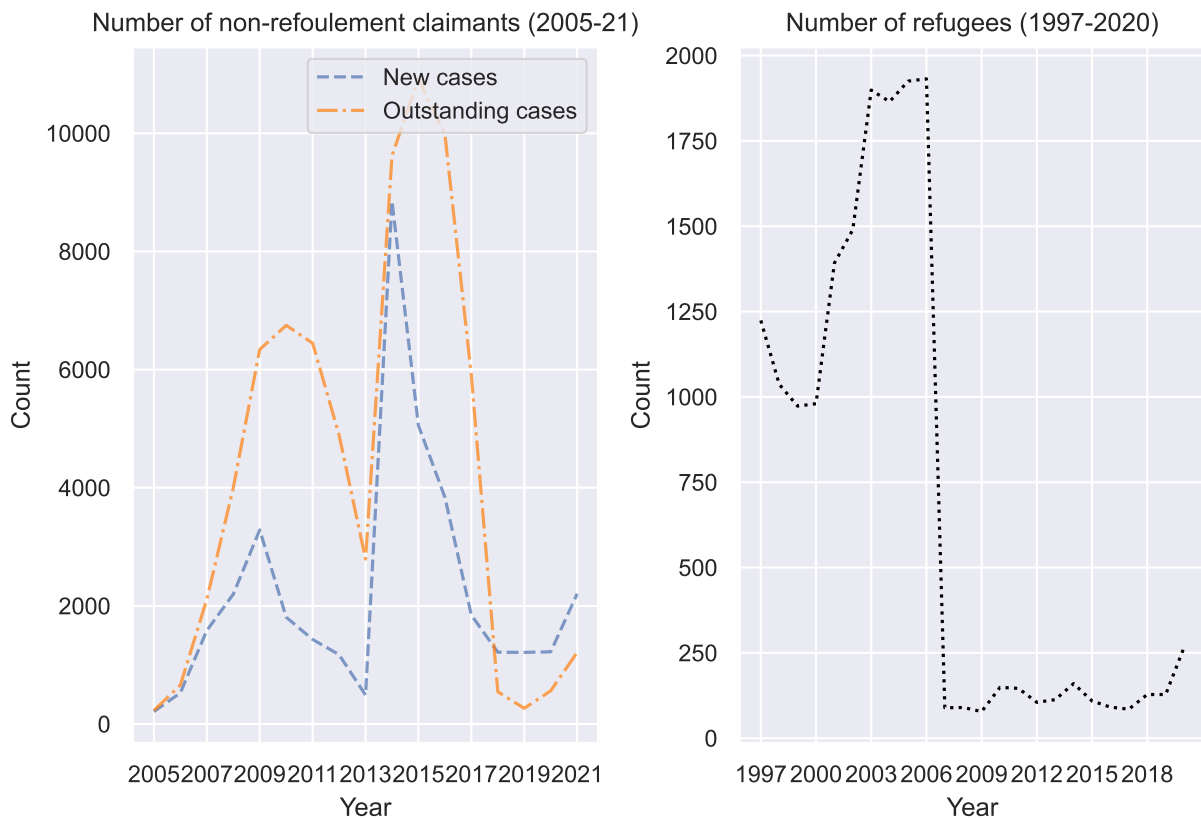
Figure 2.1: Number of non-refoulement claim cases (left) and refugees (right) in Hong Kong

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

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 (*Enforcement | Statistics on Non-*

refoulement Claim 2021) 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 (*Enforcement | Statistics on Non-refoulement Claim* 2021) . 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 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 (*Enforcement | Statistics on Non-refoulement Claim* 2021) 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.

Responses by political camps

Given the extremely small number of non-refoulement claimants, the issue of asylum seeking regime in Hong Kong hadn't gained much attention until early 2016, after a senior Immigration Department official's accusation of the USM being abused during a TV interview in October 2015 and the subsequent propagation of two fabricated WhatsApp audios alleging asylum seekers as perpetrators of robberies and homicides in the New Territories district (Ng, Choi, et al. 2019) . Soon afterwards, coverage of asylum seekers by newspaper media has increased, with two pro-Beijing newspapers named *The Sun* (which ceased publication in April 2016) and *Oriental Daily* contributing to the largest proportion of articles which depict this group of vulnerable population as a source of social problems (e.g. committing crimes) which warrant stringent measures like detention camps as solutions (Ng, Choi, et al. 2019) .

Different opinions of how the Government should deal with the asylum seekers in the political debate could also be classified according to political camps at the same time. On one side, echoing with the HKSAR Government's already restrictive policies for assessing non-refoulement
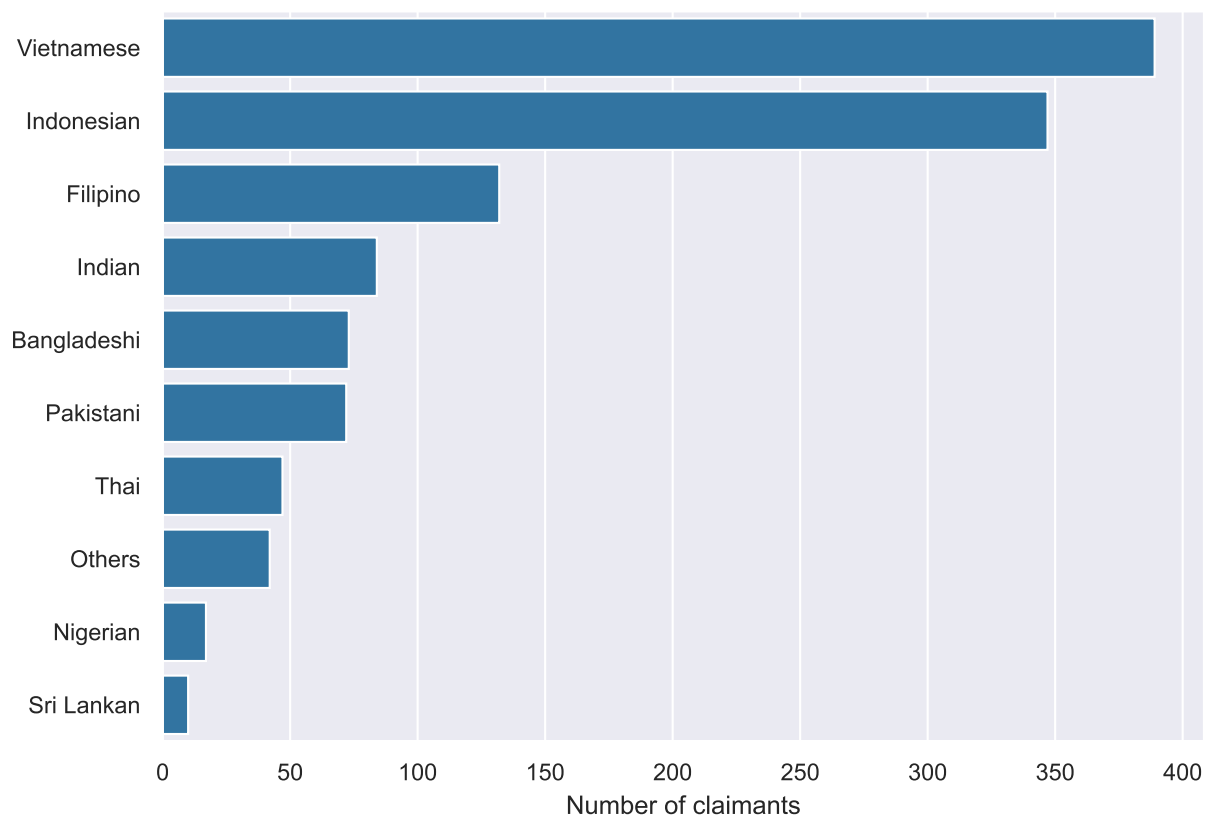
Figure 2.2: Number of non-refoulement claimants by nationality in 2021

claims, pro-Beijing camp politicians also create narratives that non-refoulement claimants and asylum seekers are anything but genuine in the legislature. For instance, less than a year after asylum seekers had gained salience in public debate, legislator Holden Chow submitted a motion in November 2016 under the banner of "combating bogus refugees" who, among other sins, were working illegally and/or committing crimes while calling for the revision of the USM and carrying out 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 DAB[1] 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 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.

By contrast, the pro-democracy camp is less hostile towards non-refoulement claimants and even advocates better treatment of these vulnerable individuals. 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 indiscriminatingly applying the proposed "draconian measures" on all asylum seekers (Cheng 2016) . 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 approval from this group of population (Justice Centre Hong Kong 2014; Vision First 2015; Refugee Union 2018) . Indeed, legislator Cheung's prominent advocacy for the rights of asylum seekers had indirectly caused him to become a victim of political smearing a few months before the 2016 Legislative Council election, as political banners imposting legislator Cheung's party were seen hanging in streets which maliciously called the legislator "the father of refugees" (Ngo 2016) . It should also be noted that, nevertheless, a

---

[1]For 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.

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 was "disturbing", he also attributed the most responsibility of "bogus non-refoulement claims" to the Government due to the latter's inefficiency and lack of knowledge in handling the issue (Cheng 2016) . 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 prosecution.

Lastly, the localist camp's stance on dealing with asylum seekers is less confrontational than that 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 on Holden Chow's anti-fake-refugees motion had two points which worth highlighting[2] (*Hui4yi4 Guo4cheng2 Zheng4shi4 Ji4lu4* 2016) (p.1232-33). Firstly, he shared the pro-democracy camp's view of the main reason for the so-called "bogus refugees" issue being the Government not allocating adequate resources to assess the claimants' applications. 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 regression in civlisation to the world. Interestingly, the localist camp's reservation of adopting more stringent measures against asylum seekers has drawn smear from some pro-Beijing medias, such as an editorial from the Sun (2016) accused the localists of "remaining silent" on non-refoulement claimants wasting public resources of Hong Kong.

---

[2]The code for tuning the model can be found in the appendix. Hyper-parameters of the tuned 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.

Chapter 3

Methods

After reviewing both the theories of the impact of national identity on xenophobia as well as how political camps in Hong Kong imagine the city's national identity, how could we know the possible associations between each camp's definition of the city's national identity and its attitude towards asylum seekers who are mainly non-ethnic Chinese? This section will discuss the strategy of answering this question.

3.1 Hypotheses

Based on the aforementioned literature in the previous section, I will formulate hypotheses to be tested and explain why I expect them to go in certain directions. The main hypothesis to be tested will be:

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

I expect the pro-Beijing camp news media to frame asylum seekers more negatively based on the literature on the relationship between national identity and xenophobia (e.g. Hjerm (1998) and Heath and Tilley (2005)) which states that a more ethnic-based conception of national identity is generally associated with a higher degree of xenophobia towards immigrants, of which asylum seekers is one of the (most vulnerable) subgroups of immigrants. Accordingly, since pro-Beijing politicians and officials (including the Chinese President *Xi Jinping* himself) (Lin and Jackson 2021) resort heavily to ethnic-based discourses for justifying Chineseness of the Hong Kong identity, pro-Beijing media outlets can be expected to hold more negative attitudes towards

⁷⁶⁷ the mostly non-ethnic Chinese non-refoulement claimants. In particular, the pro-Beijing camp's

⁷⁶⁸ emphasis on how Hong Kong citizens belong to the greater Chinese national identity based on

⁷⁶⁹ shared ethnicity with mainland Chinese might lead to its affiliated media outlets to portray

⁷⁷⁰ asylum seekers as a threat to society since their ethnic distinctiveness marks them as ever-being

⁷⁷¹ strangers of the family of the majority Chinese population as Erni (2012) describes.

⁷⁷²   By contrast, since the pro-democracy camp focuses more on civic values (e.g. civil liberties,

⁷⁷³ the rule of law) while defining the Hong Kong identity, the non-Chineseness of asylum seekers

⁷⁷⁴ should appear less threatening to the Hong Kong identity from the pro-democracy camp's point

⁷⁷⁵ of view. Lastly, the localist camp's primary focus on the mainland-Hong-Kong division regarding

⁷⁷⁶ the city's identity again prevents a clear direction for H_1 from being formulated beforehand.

⁷⁷⁷ 3.2   Data collection

⁷⁷⁸ I will study news media reports on asylum seekers and non-refoulement claimants in Hong Kong

⁷⁷⁹ by local mass media during the year of 2019. Specifically, media publishing in **Chinese** in printed

⁷⁸⁰ format will be included in the data[1]. For the retrieval of the relevant newspaper articles, I will

⁷⁸¹ rely on the *WiseNews* database which can fetch news articles with search keywords published

⁷⁸² in Hong Kong[2]. While it would have been more ideal to also include online news media into

⁷⁸³ the dataset as Lee (2018) mentions that they have become quite prominent in the 2010s for

⁷⁸⁴ providing counter-narratives to the mainstream media which are often co-opted by the Chinese

⁷⁸⁵ Government[3], WiseNews does not directly provide the texts of online media news article and thus

⁷⁸⁶ requires additional steps of web scrapping which are, unfortunately, beyond the author's capacity

⁷⁸⁷ by the time of writing. That being said, including only the printed news media during 2019

⁷⁸⁸ would still cover 16 of the well-known news media with associations to different political camps

⁷⁸⁹ according to the survey on media credibility by the Chinese University of Hong Kong (CUHK

---

[1]For 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.

[2]The code for tuning the model can be found in the appendix. Hyper-parameters of the tuned 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]The function for generating the SHAP plots for figures 4.7 - 4.10 can be found in the appendix.

3. Methods

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 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 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 permit-

ted 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 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 oppor-

tunity 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 of the limitations of using media reports as the data for this

thesis. 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. Despite these shortcomings, media reports

can still adequately help prompt the hypothesis of whether depictions of asylum seekers in Hong

Kong might be associated with their political camp's definition of the city's national identity. 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 these media reports may even serve to justify policymaking or reflect

the government's position on the issue. Drawing from Fong (2017a) and Lee (2018), the media

industry in Hong Kong is quite remarked for their political stances, and quite a number of outlets

have their owners being affiliated with the Chinese and/or HKSAR Governments. Therefore, reports published by 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 test the hypotheses proposed above.

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 media outlets to potential non-Chinese audience may also vary based on each outlet's political affiliation. 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 refer to the group of population-in-interest. 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 *outside of* Hong Kong
nan4min2	Refugee	Excluded refugees *outside of* 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 Modelling

Since the period under study spans for the year 2019, the dataset contains almost 700 articles in total which would be extremely time- and resource-consuming to qualitatively analyse each article line by line, whereas only focusing on a small subset of the articles may risk losing generalisability. On the other hand, while it is true that using human coding (whether it be done by trained experts or crowd coding platforms) for sentiment analysis will be likely to yield the most accurate results as van Atteveldt et al. (2021) have noted, doing so will be too resource-intensive for a master-level thesis, and this method may also fall short of reliability, i.e. how other researchers may replicate the study if wanted. Despite some of the caveats listed by van Atteveldt et al. (2021) (e.g. accidentally identifying spurious relationships between variables, problem of overfitting) which can be remedied with corresponding solutions, I will use supervised machine learning to test the hypothesis. This means 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.

3. Methods

Starting with the models that will be used for the analysis, I will rely on some commonly used classification models. Candidate models 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 such a model for this paper due to the relatively small size of the dataset at hand as well as the difficulty of interpreting the features' impact on the dependent variable directly from neural network models. I will first run some baseline models with the default parameter settings of the models, and then select one for further tuning based on their performance on their *f1 scores* (which measures the harmonic mean of *precision* and *recall*)[4]. Using f1 score as the evaluation metric can better measure a model's performance when there is class imbalance in the dependent variable compared to metrics like accuracy and the ROC-AUC score.

The model which obtains the overall highest f1 score will then be used for calculating the *SHAP values* of the features (Lundberg 2022) which evaluate how much impact each feature has on the model prediction when the features is at certain values versus at its baseline value. In essence, the higher the SHAP value of a feature, the higher its impact of the model's prediction.

In order 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 using a simple bag-of-word approach, I will use the term-frequency inverse-document-frequency (TF-IDF) matrix transformation which also considers in how many documents a word appears out of all the documents in the corpus. This is because words that appear too frequently across all the documents in the corpus may not convey much meaning of each doucment and thus should be treated as stop words. 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

---

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

news articles into a certain number of pre-defined latent topics (the exact number will be figured out in the ***preprocessing*** section later), and then use this matrix as part of the inputs for the models.

The majority of the implementation from data preprocessing, 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 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

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

921 that due to the medium of publication of localist news media being almost exclusively online,

922 none of the outlets in that camp was included in the dataset. Therefore, it would not be possible

923 to examine how the localist media outlets might report on asylum seekers and non-refoulement

924 claimants in news articles.

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 or the HKSAR Government)	Wen Wei Po, Hong Kong Government News, Hong Kong Commercial Daily, Ta Kung Pao
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
Localist camp	N/A
Neutral	Metro Daily, Ming Pao Daily News

925 3.4.2 Dependent variable

926 As mentioned before, the dependent variable of H_1 will be the sentiment of the news articles

927 towards asylum seekers in the polarities of *positive*, *neutral* and *negative*. But how exactly should

928 these three polarities mean in this context? Again referring to the literature can help devise a

929 sound coding scheme. Drawing McKay et al. (2011), Cooper et al. (2017) and Ng, Choi, et al.

930 (2019), *positive* articles usually acclaim the person or event in concern, describe personal stories,

931 report expert opinions without derogatory terms or depict refugees favourably, *neutral* articles

932 simply report an incident without an overt sentiment or opinion towards particular stakeholders

933 and with little interpretation by the journalists, and *negative* articles frame asylum seekers either

934 as bringing problems to the host society or use inaccurate terms to delegitimise this group of

935 population. I will follow these coding criteria for the polarities of the news articles in the dataset

936 since these definitions suit the purpose of this study. The numerical representation of the polarities

937 in the dataset will be as follows, namely, positive articles will be coded as 2, neutral articles will

938 be coded as 1, and negative articles will be coded as 0.

939 For the context of Hong Kong about the "incorrect" terms and information used for describing

940 asylum seekers, I will refer to expert opinions about whether descriptions of asylum seekers in

Hong Kong are founded[5]. 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.

3.4.3   Control variables

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`) with the first quarter beginning in January and ending in March. Since 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 how asylum seekers may be portrayed during the peak of the protest. Moreover, the presence of racial labels will be included as a control, since it is possible that more overtly racialised discourses 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

---

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

968 added to see if the lengths of the articles may be associated with the sentiment towards asylum

969 seekers.

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. In this section, I will first explore the data set, and then move onto 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.

4.1 Exploratory data analysis (EDA)

4.1.1 How does the number of news articles vary by political camps and month?

Starting with the number of articles by media outlets as shown in the left plot of figure 4.1, 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.

On a higher level of political stance, the right plot of figure 4.1 indicates that largely due to the huge volume of articles by Oriental Daily News, the pro-Beijing camp dominated the coverage of asylum seekers in Hong Kong in 2019. Meanwhile, both neutral and pro-democracy newspaper outlets published similar amounts of articles throughout 2019, and both camps constituted to
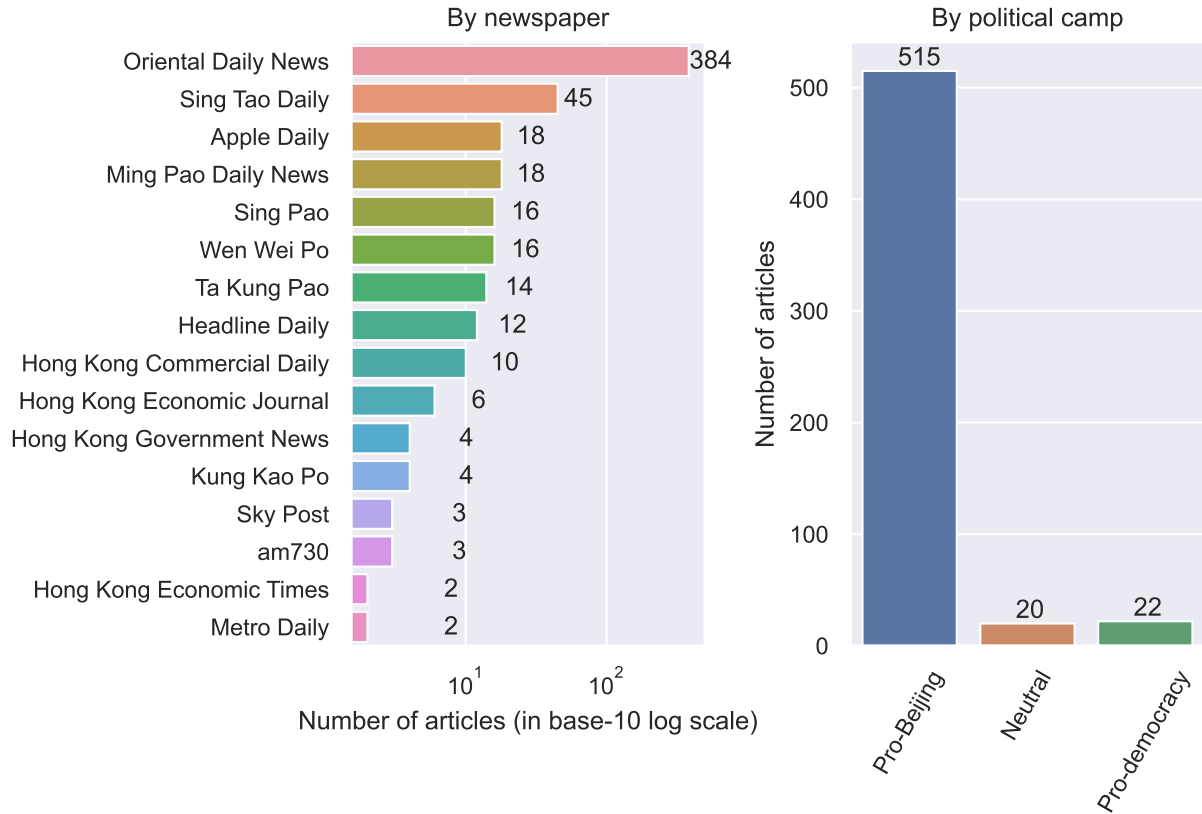
Figure 4.1: News articles on asylum seekers in 2019 by news outlet (left) and political camp (right)

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 by neutral and pro-democracy media outlets.

It will also be intriguing to see how the number of articles might vary by month in 2019. As noted before, the anti-extradition law protest lasted mostly from June to November when numerous large-scale clashes between protesters and the police occurred. From figure 4.2, it appears that coincidentally, there were the fewest amounts of 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.

In short, the majority of news articles about non-refoulement claimants in Hong Kong in 2019 were published by pro-Beijing media outlets, of which a huge proportion was from Oriental Daily News. Moreover, the number of articles by month was the lowest from August to November when

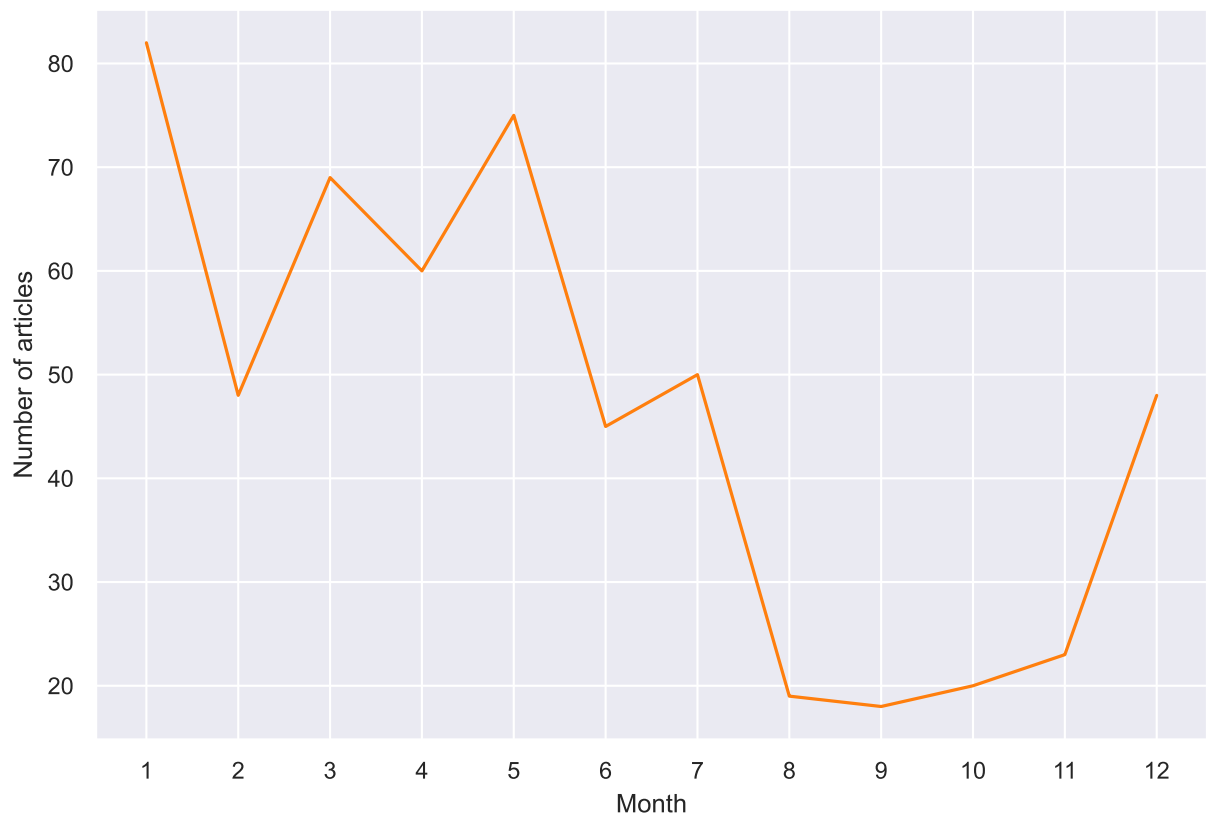**Figure 4.2:** Temporal patterns of the publication of news articles about asylum seekers in Hong Kong in 2019

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

the anti-extradition law witnessed some of the most large-scale and intense clashes.

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 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 4.3 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. In any case, the machine learning models can add the presence of racial labels as a control variable to test this potential association more formally later.

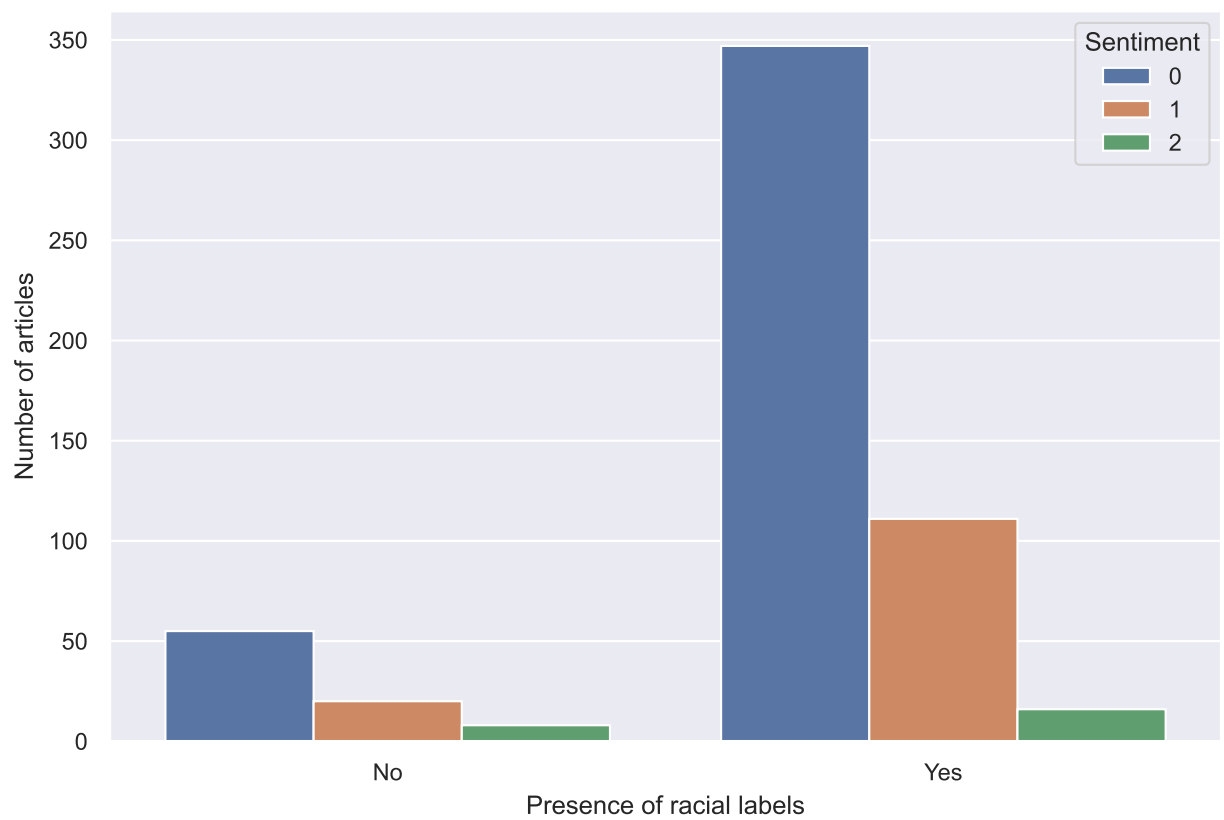**Figure 4.3:** Presence of racial labels in the news articles by sentiment
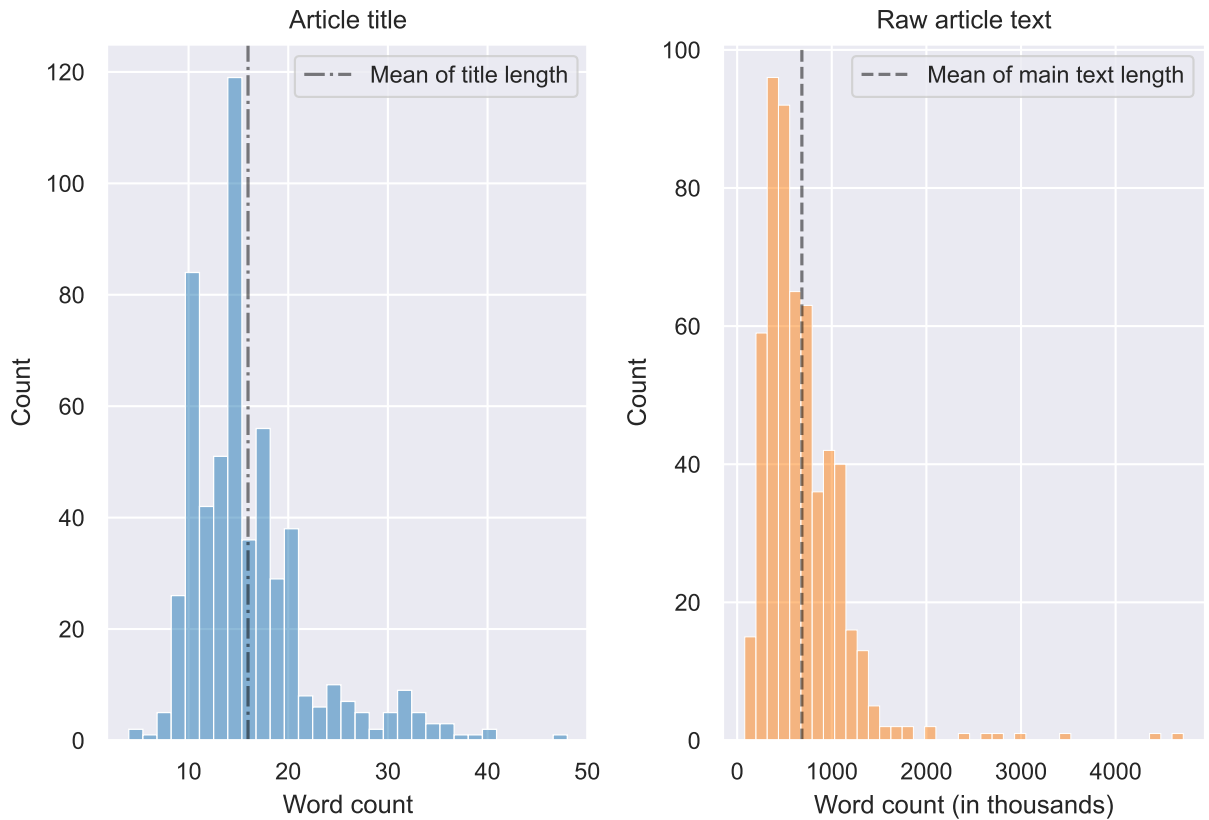
Figure 4.4: Distributions of the word counts of the articles' titles (left) and main texts (right)

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 4.4 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 of the articles.

4.2 Sentiment analysis

4.2.1 Preprocessing

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 preprocessing steps to be done so that the data are transformed into suitable formats as inputs for machine learning

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

1041 models. For starters, columns of the metadata should be excluded for being the inputs of the
1042 models. Note that I have also removed the `Newspaper` column since H_1 is more interested in
1043 whether newspaper outlets of the pro-Beijing camp *as a whole* may hold more negative attitudes
1044 towards asylum seekers in Hong Kong vis-a-vis media outlets with other political stances. The
1045 removed metadata columns are: `Index`, `Date`, `Category`, `Page_number` and `Newspaper`.

1046 Furthermore, I have binned `Month` into four even split yearly quarters (`Quarter`) to reduce the
1047 dimensionality of the dataset. A further note on the categorical features is that they will need
1048 to be transformed via one-hot encoding, meaning that each of them will be transformed into `n`
1049 variables, with `n` being the number of the original distinct values. Meanwhile, it would also be
1050 better to standardise the numerical features (i.e. other than `Political_camp` and `Quarter`) by
1051 centering theirs means at 0 for better model convergence, but the standardiser should only be
1052 fitted on the training set after splitting the data into the training and validation sets in order to
1053 avoid data leakage (the same is also true for creating the TF-IDF matrix). 20% of the observations
1054 in the dataset will be split into the test set for model validation later.

```python
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)
```

1055 The next step is to transform both the titles and main texts of the articles into a TF-IDF
1056 term-document matrix. Apart from joining the `Title` and `Text` columns together as the complete
1057 `Article`, I will also add additional words into the dictionary and remove stop words as well as
1058 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 0.493 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 sets. I set the number of latent topics (`n_components`) as 10 for the `NMF` model, and this is decided based on figure 4.5 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
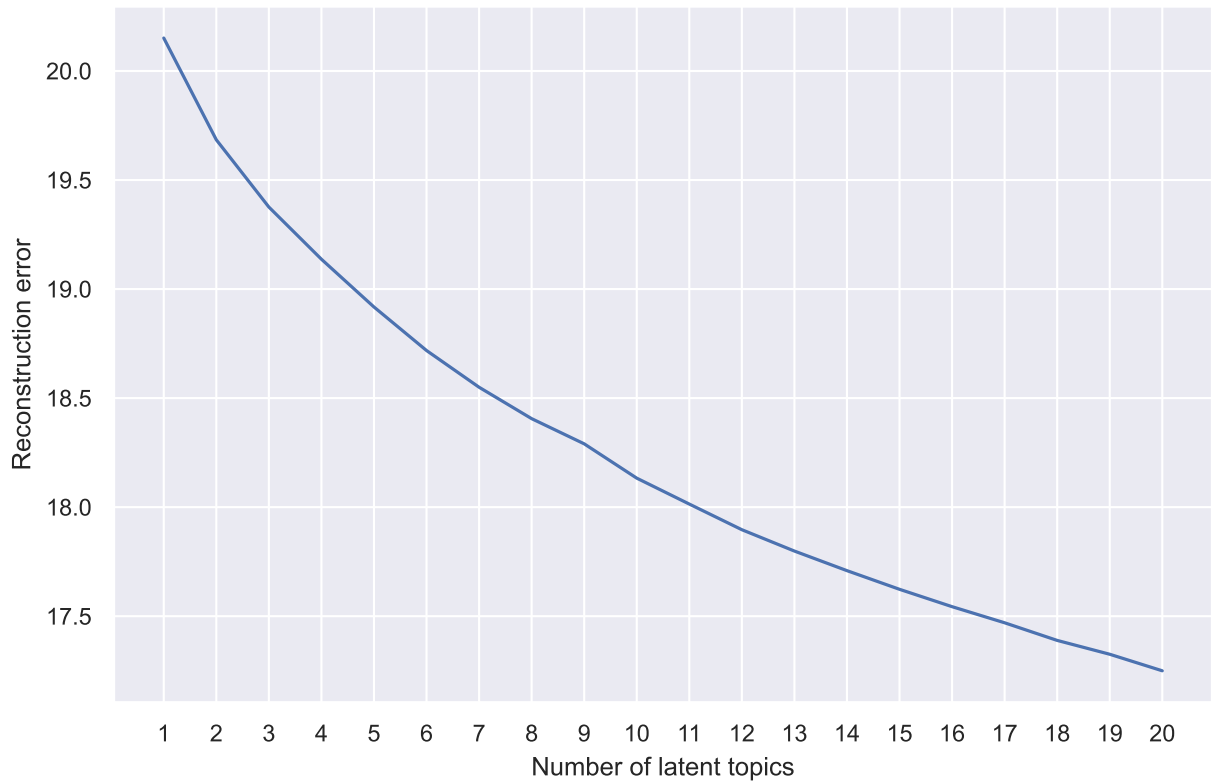
Figure 4.5: Elbow plot of the reconstruction error of NMF as a function of the number of pre-specified latent topics

¹⁰⁶⁹ appears to be a reasonable choice as a compromise between finding out a wide variety of topics

¹⁰⁷⁰ in the corpus and not fitting too much into the noise of the data.

¹⁰⁷¹   In order to make the latent topics generated by the NMF model be named more intuitive, I will

¹⁰⁷² inspect the 30 most prominent words of each latent topic (which are shown in figure 4.6) and then

¹⁰⁷³ summarise each topic. Overall, the ten topics generated by NMF are more or less semantically

¹⁰⁷⁴ coherent. Finally, I will transform the validation set's articles with the fitted instance of the

¹⁰⁷⁵ TF-IDF and NMF models on the training set. The ten latent topics that were discovered by the

¹⁰⁷⁶ NMF model are: *crimes, non-refoulement policy, illegal labours, illegal gambling, drugs, illegal*

¹⁰⁷⁷ *immigration, murder, robbery, South Asian settlements* and *public security.* The presence of the

¹⁰⁷⁸ themes in each article is represented by a value in each column of the respective theme, and the

¹⁰⁷⁹ higher the value, the more emphasised a theme is in a given article.

```
Crimes:
南亞 事主 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 盧 送院 警區 治理 追緝 逃 初步 接報 進 求助 交由

Non-refoulement policy:
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

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

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 鋪 萬元

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 職務隊 搜查 探員 追查 來源 特別 南亞 公斤 製 報

Illegal immigration:
人蛇 水警 偷渡 蛇頭 偷渡客 來港 南亞 搶灘 非法入境 西貢 截獲 警方 一般 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 裔 西分區 昨晨 籍 海里

Murder:
死者 房 劏 妻子 男子 殺 警方 單位 非洲 行街紙 死亡 同鄉 傷痕 盧 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

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

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 衞 生 街坊 搭建 渡船街 雜物 油麻地 治安 更 假難民 垃圾 吸毒

Public security:
港府 假難民 香港 政府 豉 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 4.6: 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

4.2.2 Training the model

After the above preprocessing steps, it is time to train a model that 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 and model tuning, 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 class imbalance and also set the `early_stopping_rounds` argument to `5` to prevent overfitting while training the `XGBoost` model [1]. 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, and the 5-fold cross validation f1 scores are also provided for reference. 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 both the 5-fold cross validation and test set macro average f1 score scores of the four baseline models. It seems that XGBoost and logistic regression

---

[1]For 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: Log loss on 5-fold cv and test set for the 4 baseline models

	5-fold cv f1 score	Test set 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: Comparison of the f1 scores on 5-fold cross validation and test set between the baseline and tuned XGBoost models

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

perform better out of all the baseline models. Eventually, I decided to proceed with tuning the XGBoost model because its tree-based nature allows it to capture potential non-linear relations between the features and the target variable compared to logistic regression which is a linear classifier.

Judging from the difference of the f1 scores between cross validation and test set by the XGBoost model, however, it seems that the baseline model may have under-fitted the data because the cross validation f1 score is considerably lower than the test set f1 score. I will therefore perform hyper-parameter tuning of the XGBoost model to see if it can be fitted better for the training data[2]. The comparison of the baseline and tuned model's performance in f1 score can be found in table 4.4. It seems that the tuned model now fits the training data much better because the f1 score on cross validation is very close to that on the test data. Table 4.5 further breaks down the f1 score of the tuned xgboost model on the testing data. Although the model performs relatively worse in predicting news articles with neutral sentiments since the f1 score for this class is only about 0.69, it performs quite well on predicting the classes of positive and negative since these two classes' per-class f1 score are both close to 0.9. I will thus calculate the SHAP values of the features with the tuned model given its better performance overall on fitting the training data and predicting unseen test data.

---

[2]The code for tuning the model can be found in the appendix. Hyper-parameters of the tuned 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.

Table 4.5: Classification report on the f1 score of the tuned XGBoost model on the testing 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 Is the pro-Beijing camp more likely to portray asylum seekers in 2019 more negatively than other outlets?

This section will move onto 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 intriguing findings afterwards.

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[3]. According to figure 4.7, 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 fourth most important features in predicting the sentiment of a news article, and its magnitude of SHAP value in affecting the model's output is only lower than those of three news article themes about asylum seekers (i.e., public security, murder and 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 considerably more informative for predicting whether an article has positive polarity (`Class 2` in blue) or not and has negative polarity (`Class 0` in pink) or not, but no so much while predicting if an article has a neutral polarity (`Class 1` in green) or not. Therefore, it seems that pro-Beijing media outlets do tend to differ their sentiment 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

---

[3]The function for generating the SHAP plots for figures 4.7 - 4.10 can be found in the appendix.
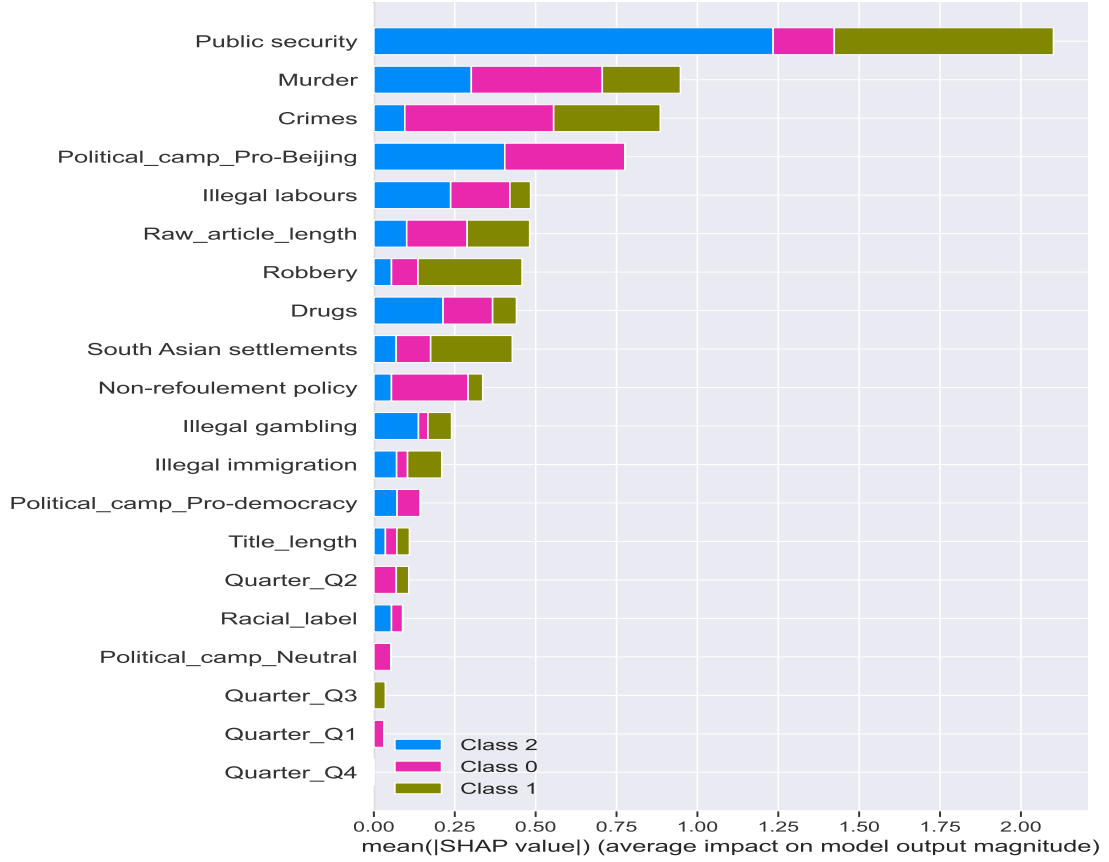
Figure 4.7: Magnitudes of SHAP values of each feature for all classes

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 categories
may to some extent be used interchangeably, or the media find it difficult to not mention both
at the same time 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 titles of news articles.

Nevertheless, figure 4.7 does not really show the *direction* of the SHAP values of each feature
in affecting the model's output. We can therefore use the beeswarm 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. Not that the importance of the features are arranged in descending
order on the y-axis from top (the most important) to bottom (the least important). In a SHAP
value beeswarm plot, dots in red 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 blue mean the value of a feature
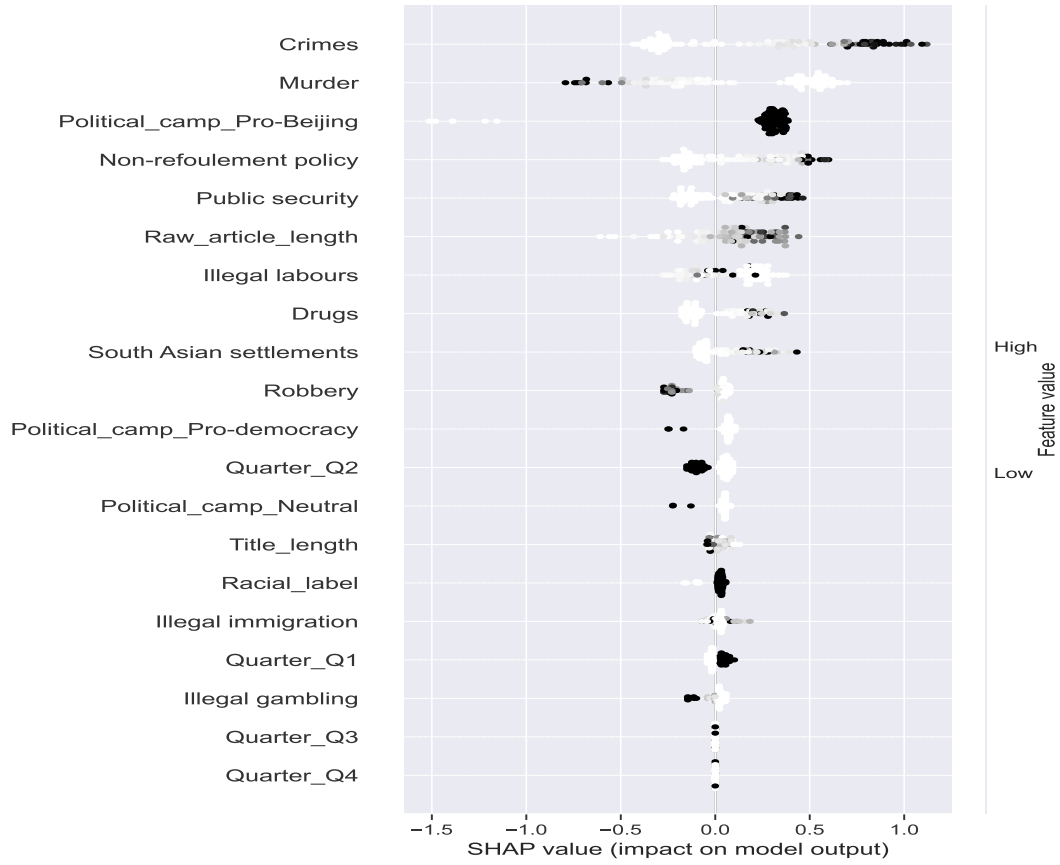
Figure 4.8: The SHAP values of the features in the prediction of whether an article has a negative polarity

is low (or absent in the case of a binary feature). Moreover, a dependence plot which zooms in the relations between the feature values and SHAP values of the eight most important features will be provided.

4.3.2 SHAP values of predicting negative news articles

According to figure 4.8, 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 magnitude 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,
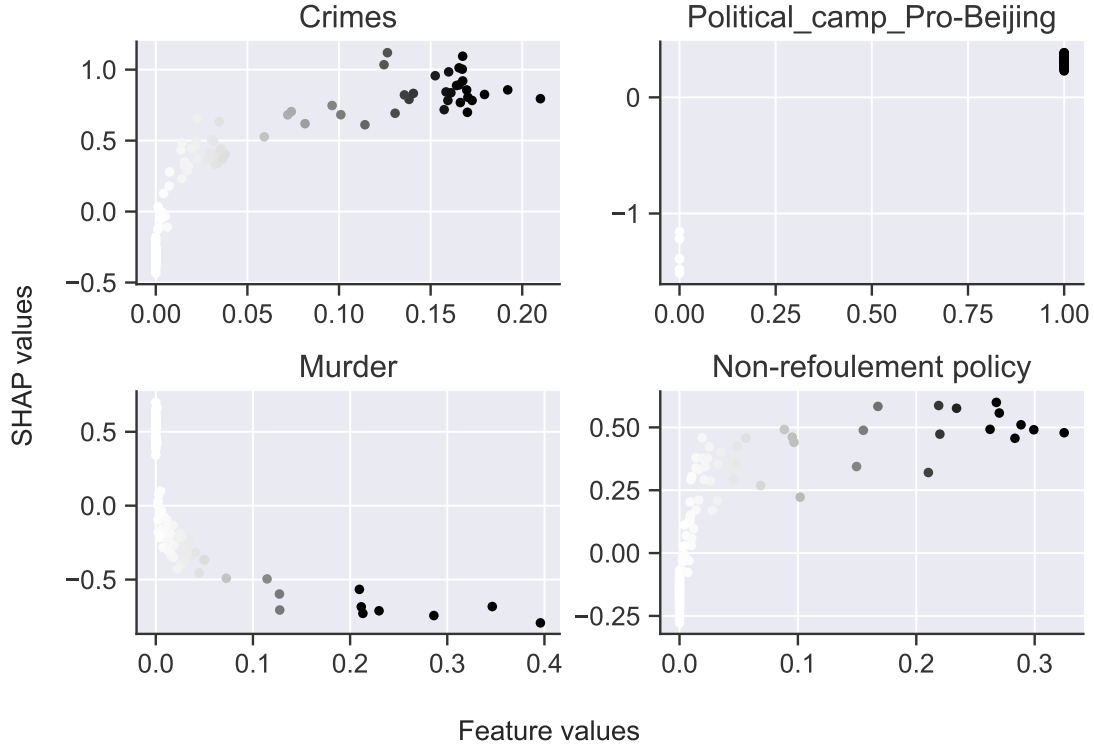
Figure 4.9: Dependence plot of the eight most important features for predicting negative polarity of news articles

meaning that pro-Beijing camp newspapers are more likely to depict non-refoulement claimants negatively than their counterparts in other political stances[4].

Figures 4.9 and 4.10 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 here. Firstly, articles which are more related to crimes (`Crimes`), the policies of non-refoulement claims (`Non-refoulement policy`) and public security (`Public security`) are more likely to report on asylum seekers negatively. Apart from the red dots in each subplot which indicate higher presence of these themes in the articles, these themes already have positive SHAP values (i.e. pulling the predictions towards the negative polarity) even when their feature values are relatively low. These directions are consistent with the most prominent words found in these topics. Specifically, both `Non-refoulement policy` and `Public security` have the derogatory term "fake refugee" as one of the topic words, whereas the theme `Crimes` found the noun phrase "South Asian army" (*nan2ya4 bing1tuan2*) as two of the topic words which together attempt to

---

[4]Of course, Oriental Daily News contributed to a huge volume of reportage on non-refoulement claimanats within the pro-Beijing camp. I will explain why it is not appropriate to re-run the model without entries from this outlet in the Discussion section below.
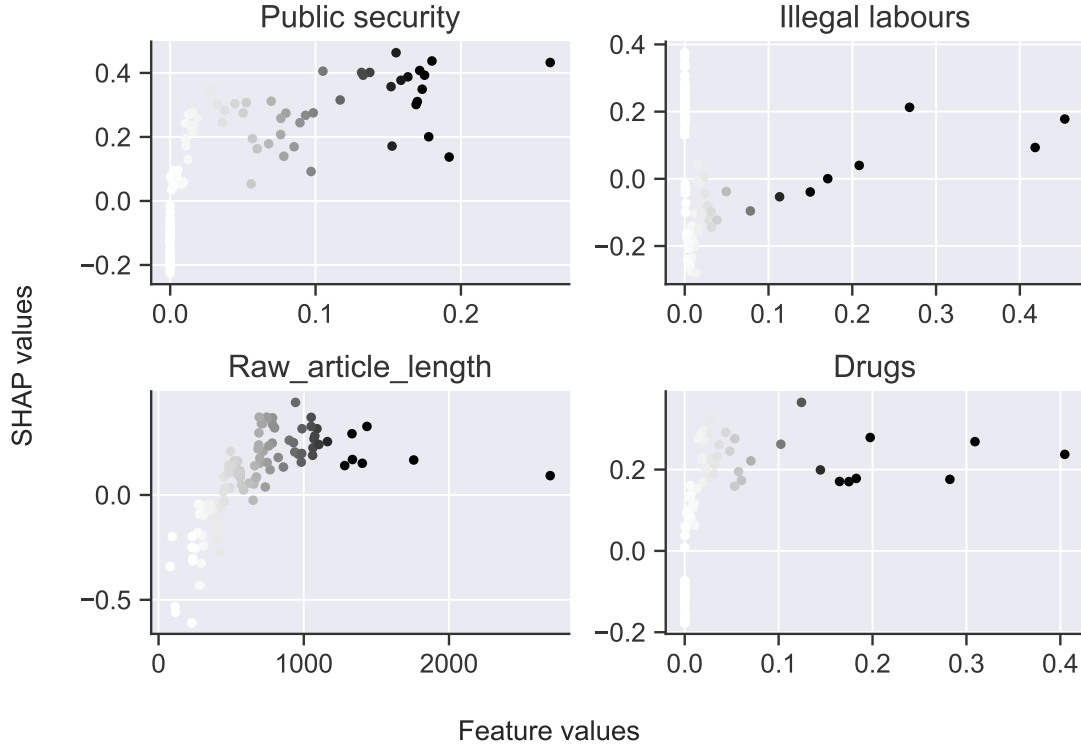
Figure 4.10: Dependence plot of the eight most important features for predicting negative polarity of news articles

conflate non-refoulement claimants with the image of them being South Asians coming to Hong Kong en masse for committing crimes. Also, the relationship between the feature and SHAP values of these features appear to be non-linear.

Interestingly, articles with longer passages are also predicted to be more likely for holding negative opinion towards non-refoulement claimants (`Raw_article_length`).

4.3.3   SHAP values of predicting neutral news articles

Figure 4.11 shows the SHAP values of each feature in contributing to the prediction of neutral news articles 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. Rather, some topics of the news articles are more informative in predicting whether news articles are only reporting incidents without interpretations from the journalists. We can look into this further in figures 4.12 and 4.13.
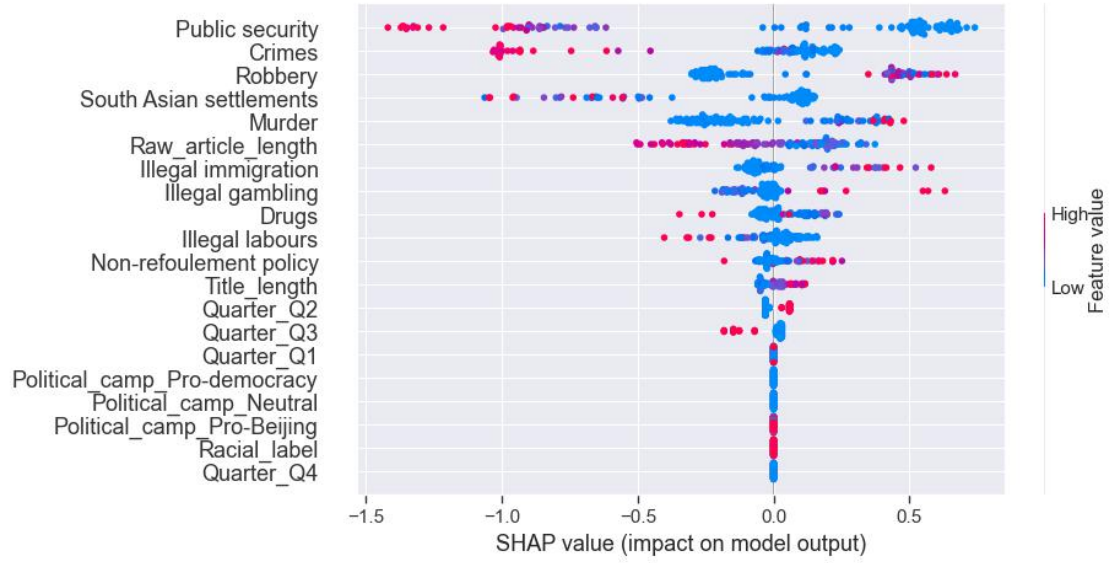
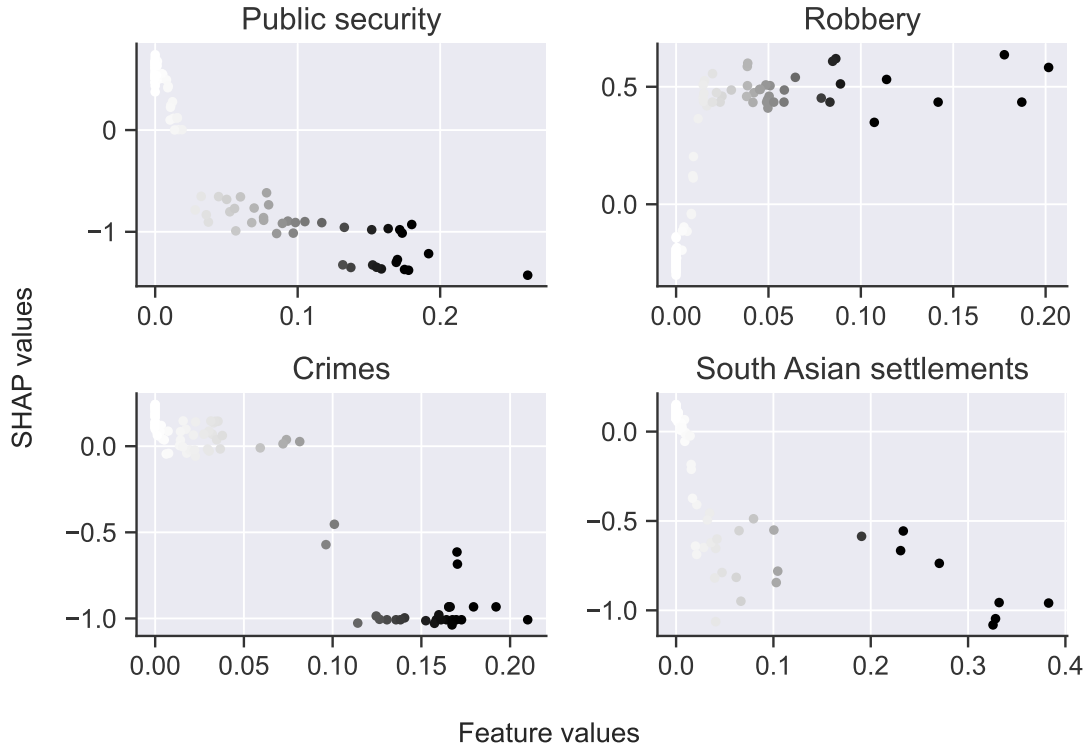**Figure 4.11:** The SHAP values of the features in the prediction of whether an article has a neutral polarity

Figure 4.12: Dependence plot of the eight most important features for predicting neutral polarity of news articles
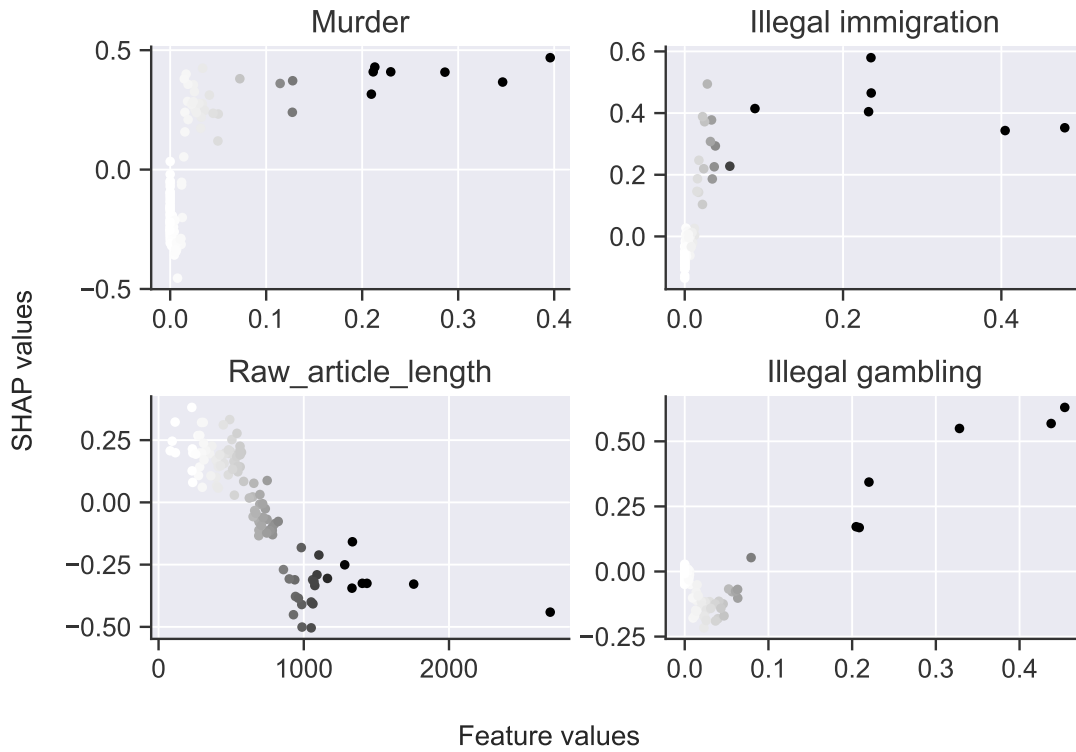
Figure 4.13: Dependence plot of the eight most important features for predicting neutral polarity of news articles

For instance, articles more related to `Public security` or `Crimes` 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

On the other hand, articles related to `Robbery` were generally covered more neutrally by newspapers. Moreover, as the length of the news articles gets longer, it becomes less likely that the articles are not loaded with conspicuous sentiments towards asylum seekers in Hong Kong.

4.3.4   SHAP values for predicting positive news articles

Lastly, figure 4.14 shows the SHAP values of the feature in predicting whether news articles have a positive polarity towards non-refoulement claimants or not. This time, pro-Beijing affiliation becomes the second most crucial feature for predicting

4.4   Discussion

In general, the above results are consistent with those found by Ng, Choi, et al. (2019) on articles about asylum seekers back in when the issue first gained its salience in 2016. Although this thesis

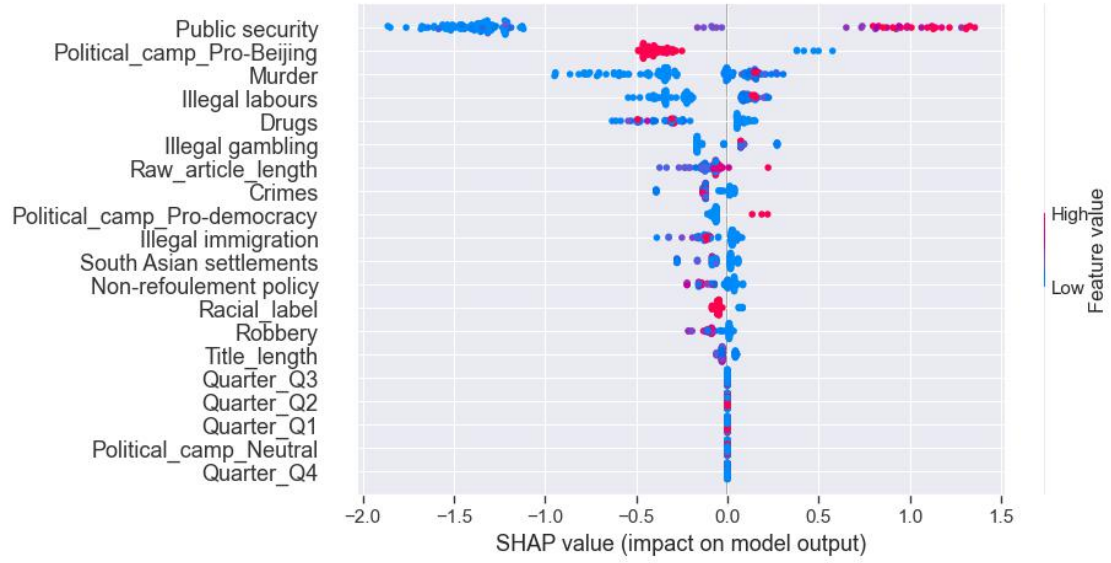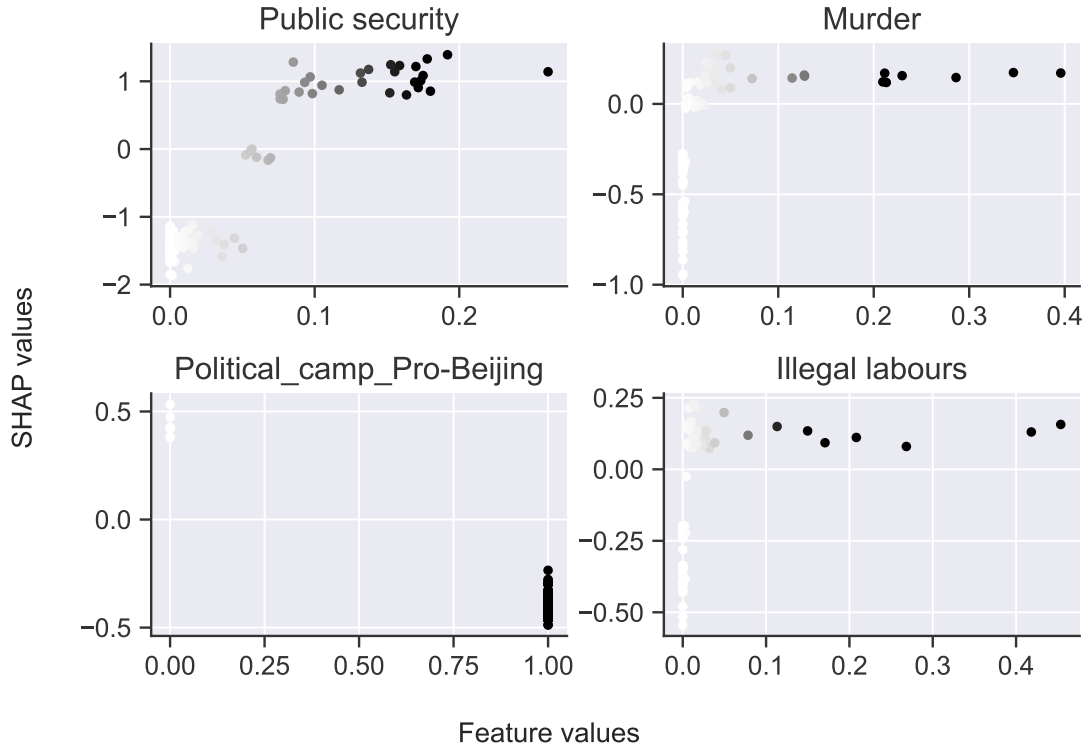**Figure 4.14:** The SHAP values of the features in the prediction of whether an article has a positive polarity

Figure 4.15: Dependence plot of the eight most important features for predicting positive polarity of news articles
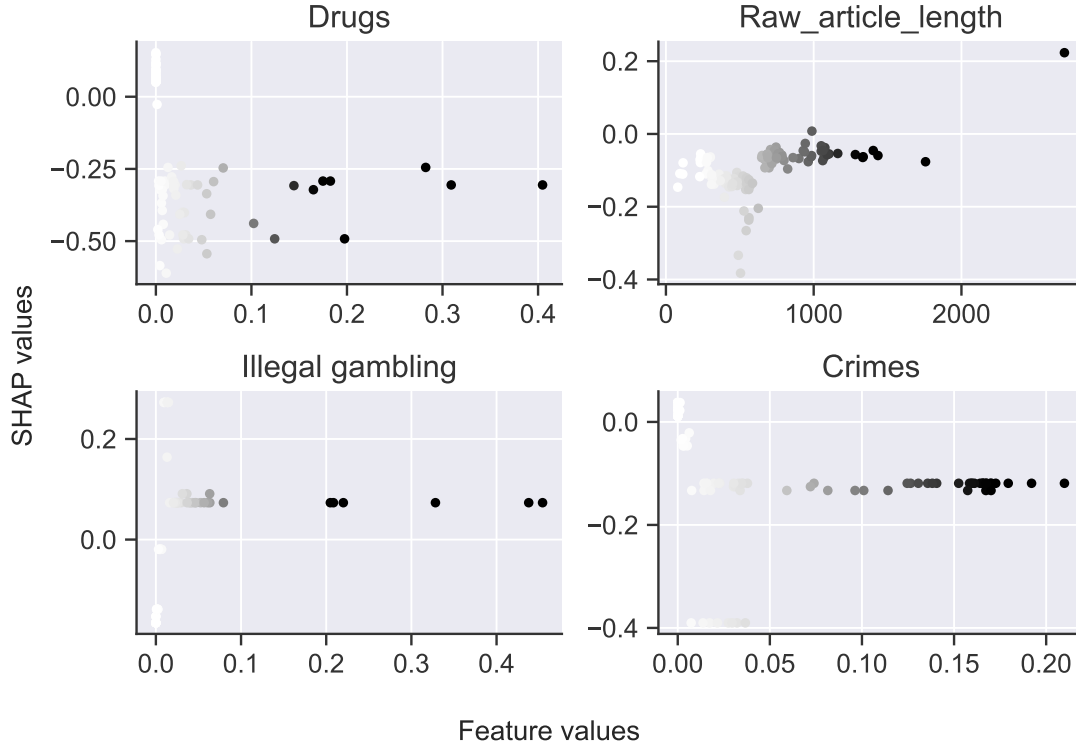
Figure 4.16: Dependence plot of the eight most important features for predicting positive polarity of news articles

focuses less on *how* non-refoulement claimants are portrayed in the articles by the media,

One question about the validity of the analysis result can be that the vast volume of publication about asylum seekers by Oriental Daily News might have skewed the pro-Beijing media's attitudes towards asylum seekers towards the negative ends. While it is certainly a fact that Oriental Daily News covered non-refoulement claimants disproportionately throughout 2019, it would not be ideal to re-run the model after excluding the news entries by this outlet for two reasons. Firstly, Oriental Daily News is after all part of the pro-Beijing media network (and in the context of reportage on asylum seekers within the city, a very important one) whose owners are co-opted by the Beijing Government according to Lee (2018), and thus dropping out this particular newspaper will lose a lot of information about how the pro-Beijing overall may portray asylum seekers. Secondly, the fact that Oriental Daily News had the second largest market share among other paid newspapers and was the third most popular media by 2019 (Adintime 2021) implies that dropping out this media outlet

However,

Chapter 5

Conclusion

In conclusion, it is found that at least in 2019, the political camp of media outlets was associated with their attitudes towards asylum seekers in Hong Kong. Specifically,

- I will sum up the findings after editing the main sections

5.1 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 2016), such bills will very unlikely be opposed by others, and the voting results

5. Conclusion

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 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 certain issues, including on the status and legitimacy of non-refoulement claimants. As the study by Boomgaarden and Vliegenthart (2009) shows, the tone of how asylum seekers are reported in news media is negatively associated with the public attitude towards asylum seekers in later periods, meaning that more pejorative coverage of refugees by the media may cause the public to be more averse towards this group of vulnerable population, 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, the gradual decrease of

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.

```
import numpy as np

np.random.seed(1)
```

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 stocastic nature of the search space defined for Bayesian Optimisation.

```python
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)
```

```python
  }


# Defining the objective function
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_
```

References

```python
#   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")

#   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 (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.

Börzel, Tanja A. and Thomas Risse (2018). "From the Euro to the Schengen Crises: European
Integration Theories, Politicization, and Identity Politics". In: *Journal of European Public
Policy* 25.1, pp. 83–108.

Chen, Tianqi et al. (2015). "Xgboost: Extreme Gradient Boosting". In: *R package version 0.4-2*
1.4, pp. 1–4.

Cheng, Kris (Dec. 2016). *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/.

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/.

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.

References

1298 Cooper, Samantha et al. (2017). "Media Coverage of Refugees and Asylum Seekers in Regional

1299   Australia: A Critical Discourse Analysis". In: *Media International Australia* 162.1, pp. 78–89.

1300 CUHK (2020). *Public Evaluation on Media Credibility.*

1301   https://ccpos.com.cuhk.edu.hk/public-evaluation-on-media-credibility/.

1302 elgarteo (2021). *HKPolDict: Hong Kong Politics Dictionary for Text Mining.*

1303   https://github.com/elgarteo/HKPolDict.

1304 *Enforcement | Statistics on Non-refoulement Claim* (2021).

1305   https://www.immd.gov.hk/eng/facts/enforcement.html.

1306 Erni, John Nguyet (2012). "Who Needs Strangers? Un-imagining Hong Kong Chineseness". In:

1307   *Chinese Journal of Communication* 5.1, pp. 78–87.

1308 Fong, Brian (2017a). "In-between Liberal Authoritarianism and Electoral Authoritarianism:

1309   Hong Kong's Democratization under Chinese Sovereignty, 19972016". In: *Democratization*

1310   24.4, pp. 724–750. ISSN: 1351-0347, 1743-890X. DOI: 10.1080/13510347.2016.1232249.

1311 Fong, Brian CH (2017b). "One Country, Two Nationalisms: Center-periphery Relations between

1312   Mainland China and Hong Kong, 19972016". In: *Modern China* 43.5, pp. 523–556.

1313 *Full Text of Constitution of Communist Party of China - Xinhua* (Nov. 2017).

1314   http://www.xinhuanet.com/english/special/2017-11/03/c_136725945.htm.

1315 Greenfeld, Liah and Jonathan Eastwood (2007). "Nationalism and National Identity:

1316   Definitional Issues". In: *The Oxford Handbook of Comparative Politics.* Ed. by Carles Boix

1317   and Susan Stokes. 1st. Oxford University Press, pp. 256–273.

1318 Heath, Anthony F. and James R. Tilley (2005). "British National Identity and Attitudes

1319   towards Immigration". In: *International Journal on Multicultural Societies* 7.2, pp. 119–132.

1320 Hjerm, Mikael (1998). "National Identity: A Comparison of Sweden, Germany and Australia".

1321   In: *Journal of Ethnic and Migration Studies* 24.3, pp. 451–469.

1322 Home_Affairs_Department (Jan. 2018). *The Demographics: Ethnic Groups.*

1323   https://www.had.gov.hk/rru/english/info/info_dem.html.

1324 "Hong Kong" (Jan. 2022). "Hong Kong: Pro-democracy Website Citizen News Is Latest to

1325   Close". In: *BBC News.*

References

1326   *Hui4yi4 Guo4cheng2 Zheng4shi4 Ji4lu4* (Dec. 2016).

1327   https://webcache.googleusercontent.com/search?q=cache:zyvAUu3wlSoJ:https://www.legco.gov.hk/yr16-

1328   17/chinese/counmtg/hansard/cm20161201-translate-c.pdf+&cd=8&hl=zh-

1329   TW&ct=clnk&gl=ch.

1330   International, Amnesty (2022). *Refugees, Asylum-seekers and Migrants.*

1331   https://www.amnesty.org/en/what-we-do/refugees-asylum-seekers-and-migrants/. Amnesty

1332   International.

1333   Junyi, Sun (Dec. 2021). *Jieba.*

1334   Kaeding, Malte Philipp (2017). "The Rise of" Localism" in Hong Kong". In: *Journal of*

1335   *Democracy* 28.1, pp. 157–171.

1336   Kwan, Justin P. (2016). "The Rise of Civic Nationalism: Shifting Identities in Hong Kong and

1337   Taiwan". In: *Contemporary Chinese Political Economy and Strategic Relations: An*

1338   *International Journal* 2.2, pp. 941–973.

1339   Kwong, Ying-ho (2016). "The Growth of "Localism" in Hong Kong. A New Path for the

1340   Democracy Movement?" In: *China Perspectives* 2016.2016/3, pp. 63–68.

1341   Lau, Jessie and Shui-yin Sharon Yam (Dec. 2021). *'Patriots Only': Hong Kong's New Election*

1342   *System in Action.*

1343   https://thediplomat.com/2021/11/patriots-only-hong-kongs-new-election-system-in-action/.

1344   Lau, Tracy (2013). "State Formation and Education in Hong Kong: Pro-Beijing Schools and

1345   National Education". In: *Asian Survey* 53.4, pp. 728–753.

1346   Laver, Michael, Kenneth Benoit, and John Garry (2003). "Extracting Policy Positions from

1347   Political Texts Using Words as Data". In: *American political science review* 97.2,

1348   pp. 311–331.

1349   Law, Kam-Yee and Kim-Ming Lee (Mar. 2012). "The Myth of Multiculturalism in 'Asia's World

1350   City': Incomprehensive Policies for Ethnic Minorities in Hong Kong". In: *Journal of Asian*

1351   *Public Policy* 5.1, pp. 117–134. ISSN: 1751-6234. DOI: `10.1080/17516234.2012.662353`.

1352   Lecours, André (2000). "Ethnic and Civic Nationalism: Towards a New Dimension". In: *Space*

1353   *and Polity* 4.2, pp. 153–166.

1354   Lee, Francis LF (2018). "Changing Political Economy of the Hong Kong Media". In: *China*

1355   *Perspectives* 2018.2018/3, pp. 9–18.

References

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 (Jan. 2022). *Slundberg/Shap*.

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.

Making a Claim for Non-refoulement Protection in Hong Kong (2021). https://www.immd.gov.hk/eng/useful_information/non-refoulement-making-claim.html.

Manifesto (2021). https://www.dphk.org/index.php?route=eng/about_us/detail&id=1.

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 â€�“Boat Peopleâ€™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.

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.

References

Ng, Isabella, Sharice Fungyee Choi, and Alex Lihshing Chan (May 2019). "Framing the Issue of Asylum Seekers and Refugees for Tougher Refugee Policyâ€"a Study of the Mediaâ€™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.

Noury, Abdul and Gerard Roland (2020). "Identity Politics and Populism in Europe". In: *Annual Review of Political Science* 23, pp. 421–439.

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.

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.vfnow.org/7401/vision-first-supports-lawmaker-fernando-cheung-chiu-hung/.

Refugee Population by Country or Territory of Asylum - Hong Kong SAR, China | Data (2021). https://data.worldbank.org/indicator/SM.POP.REFG?locations=HK.

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.

References

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.

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/.