Open Journal of Social Sciences, 2023, 11, 18-31
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

A Study on Diversity Prediction with Machine Learning and Small Data

Rezza Moieni, Peter Mousaferiadis, Leyla Roohi

Cultural Infusion Pty, Ltd., Melbourne, Australia
Email: Rezza.m@culturalinfusion.org.au

How to cite this paper: Moieni, R., Mousaferiadis, P., & Roohi, L. (2023). A Study on Diversity Prediction with Machine Learning and Small Data. *Open Journal of Social Sciences, 11,* 18-31.
https://doi.org/10.4236/jss.2023.112002

Received: November 30, 2022
Accepted: February 6, 2023
Published: February 9, 2023

Abstract

There are discussions about the importance of diversity in literature and in the media and minimizing gaps between minorities and majorities. In order to see if a community is making progress in minimizing these gaps and to measure success, there is an interest in being able to predict the diversity of communities given currently prevailing. There are well-designed data forecasting algorithms in data science using large data sets. However, diversity data has only been collected over the last few decades. This paper adopts algorithms formulated by Grey and ARIMA (Auto-Regressive Integrated Moving Average), using small data to predict the likely diversity of a cohort for a time in the near future. Our results demonstrate there is more confident forecasting for "country of birth", but in terms of predicting linguistic and religious diversity, due to the changeable nature of these factors throughout an individual's life, we would require further data to make any accurate prediction.

Keywords

Diversity, Data Forecasting, Small Data, Mutuality, Diversity Atlas

1. Introduction

Globalization has increased communication among people with diverse backgrounds, beliefs, and cultures around the world. These interactions have become more frequent and occur more rapidly in diversely populated nations. Individuals demonstrate a preference to be a part of the international economy in both their personal and professional lives. People are also facing and accepting many challenges and significant competition from different regions of the world. It is therefore very important to discuss the diversity of a cohort to maximize and capitalize on different elements of diversity such as religion, ethnicity, language and other demographic factors including gender, age, education level, disability and sexuality. The management of diversity within organisations has increasingly been a subject of wide debate and discussion. In this current social climate, managers and associated staff should be skilled enough to prepare themselves to learn and train on multicultural ethics and differences within their customer base and in their own organisational makeup. The increasingly pronounced cultural differences between individuals within organisations ensure that diversity is critically important to discuss amongst stakeholders on every level of society (Cox, 1991). Globally, organisations are annually spending billions of dollars on training and education about diverse cultures and their predictive outputs. Demographic prediction is already widely used in marketing circles. The reason for this, fundamentally, is that this expenditure of time, energy and money has the potential to create a better business hub and marketplace where employees from diverse cultures and nationalities work together and perform best to lift the international economy. Not only in the business world but educational institutions and governments of various countries are now focusing on selecting students from different cultures and enrolling them in universities to get maximum talent from diverse populations (Page, 2007). The value of diversity and its benefits is both logical and very easy to understand. However, it has many associated complications and complexities if not managed carefully. Under a properly managed framework of diversity, people can potentially improve their abilities and interpersonal skills. Eventually, they will meet their ultimate potential, thus benefiting their organisations and educational institutions. The primary benefit of diversity in organisations is to provide a pillar based on differences in ideas and cultural values. The flow-on effect is the construction of a strong economy. Some action plans and strategies will reflect the article's main idea with the help of authentic research studies and approved theories and practices.

The impact of cultural diversity on social cohesion and governance has also been studied by (Baldwin & Huber, 2010; Helpas, 2022). (Helpas, 2022) pointed out that many believe that cultural diversity in a country has a negative impact on governance and social cohesion and then tests this theory, not only failing to prove it but also revealing that in long-established democracies cultural diversity is better accepted. It would certainly be interesting to test the widely accepted hypotheses, that cultural diversity is costly, in terms of institutional and economic performance, human development and generalized interpersonal trust cultural diversity with the scores of governance quality, economic performance in terms of competitiveness, human development and generalized trust in each country.

While there have been attempts to predict the population of a country, such as Australian Bureau of Statistics (ABS) Population Projections, there is no previous research on predicting the ethnolinguistic diversity of a community, using data science methods based on the previous and current trends

The inter-dependency between different pillars of cultural diversity has been reviewed by (Brubaker, 2013; Krech, 2013). Language and worldviews are the two most socially and politically consequential domains of cultural difference in most countries. (Brubaker, 2013) begins by aligning language and religion, provisionally, with ethnicity and nationhood.

The Project objective is to research the application of applied mathematics and artificial intelligence to the quantification of multi-factor cultural dimensions, with particular regard to the predictability of future performance.

The ability of government and businesses to use this research within their planning and budgeting processes, however, has been constrained by inadequate data-sets. Our research proposed to develop a machine learning model that would allow predictive modelling based on smaller and fragmentary data-sets.

Our previous research on this project has already solved the initial problems of doing so. The previous phase of the research worked towards the mathematical predictability of correlated factors.

This research was conducted in the following phases:

- Literature review and preliminary experimentation to establish the most promising paths for development;
- Data assemblage (Net Overseas Migration (NOM) data and short series, incomplete Census data);
- Data cleaning and normalising;
- Analysis of data to develop machine learning models (NOM Data);
- Testing data in machine learning models (Census data).

1.1. Cultural Benefits of Diversity Prediction

Predictive modelling is a statistical technique to predict future behaviour. Predictive modelling solutions are a form of data-mining technology that works by analysing historical and current data and generating a model to help predict future outcomes. It is commonly used by business and government to plan. In predictive modelling, data is collected, a statistical model is formulated, predictions are made, and the business plan is validated or revised as additional data becomes available.

We cannot talk about diversity prediction without considering the impact of overseas migration to Australia and there is good data around this parameter dating all the way back to 1900, giving us a good base from which to forecast future trends. This, therefore, will be the key parameter under consideration in our development of an algorithm to predict the diversity of an organisation, based on the four categories of country of birth, language, religion and ethnic background. Predictive analyses will be based on the history of these pillars in the organisation, in consideration of the Net Overseas Migration, when other parameters are fixed. NOM (Net Overseas Migration) data for Australia has been demonstrated in **Figure 1**.

It can be seen that the NOM has an overall upward trend with recognisable points of intervention in the 1920s and 2008-09 and no seasonal patterns. Statistical tools tell us that there is an 86% positive auto correlation in the data. Thus, the migration trend at any given year is strongly explained by the previous year's

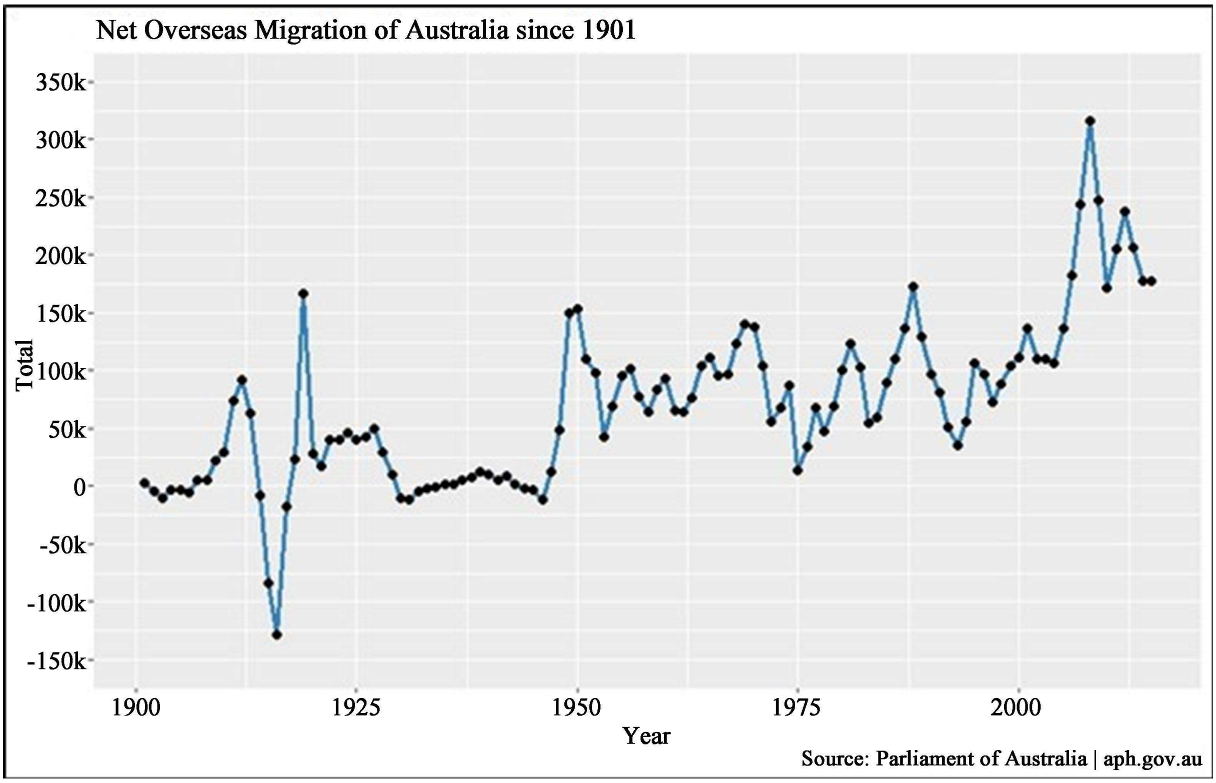

Figure 1. Net overseas Migration (Australia).

trend, and the overall NOM is most likely to show a similar pattern in coming years. The time series analysis on the data carried out using methods suggested by (Cryer & Chan, 2008) suggests that the NOM will hover around 220 k to 230 k up until 2025.

The cultural background of the current resident population is also a key influencer and must be taken into account. Until recently, the United Kingdom (UK) has always been the primary source for permanent migration to Australia. However, there are other countries such as China, India and New Zealand which have started to provide higher numbers of permanent migrants. Changes in the resident population of Australia over time can be seen here:

It is now becoming increasingly important for a broader range of organisations, when planning, to consider how best to serve future populations, customers, constituents and customers. In particular, if the diversity of the community is not considered by governments when planning for essential services and facilities, then it is obvious that the appropriateness and effectiveness of these assets will not be maximised, and indeed may miss the mark entirely. Predictive technology where cultural diversity and inclusion are concerned therefore becomes crucial to effective public policy.

Predictive modelling was not previously possible for diversity, because the available data-sets were either small, fragmentary (for example, did not include important factors), or only covered short time series. Both of these problems also precluded applying machine learning to predictive modelling, as machine learning requires large data-sets. This problem affects government budgeting and planning at all levels, and business planning and marketing.

Understanding cultural diversity in mixed communities and its impacts on groups and the wider population is very complex and has many associated challenges. A complete and comprehensive research body, therefore, is required to address the questions circulating about the pros and cons of cultural diversity (Jonsen et al., 2011). Some key benefits of cultural diversity in corporations and organisations are that it provides out-of-the-box thinking to provide solutions to problems, can foster better business productivity, and discovery of innovative ideas to grow the organisations. It suggested that organisations should develop their competencies to get maximum benefits from cultural diversity in three dimensions; Inclusion, Diversity, and Mutuality. Existing technology does not support the measurement of these characteristics (Lorenzo et al., 2017; Moieni & Mousaferiadis, 2022). Customer satisfaction dependent on providing good services to its premium clients, can only be achieved if these three dimensions are addressed. Companies should strive to get customer satisfaction by providing optimal assistance. Cultural diversity can be assessed with the help of recent research showing that customers who are of a particular nationality trust people they perceive as being like themselves. In addition, diverse language skills and cultural environment might fulfill customers' requirements more precisely. Hence, cultural diversity in organisations should be directly proportional to the multicultural society.

1.2. Benefits of Predicting Diversity in Workplaces

Diversity in the workplace represents a holistic situation. Differences among people exist in many organisations. People may behave differently in a team when other people around are from different ethnolinguistic backgrounds. Sometimes these variations increase in complexity based on cultural background, religious beliefs or psychological aspects like gender identity and worldview (Cletus et al., 2018). Studies have shown that workplace diversity enhances the employees' professional skills, improves critical thinking and the quality of problem solving (Green et al., 2022). In addition, diverse and inclusive workplaces contribute to the company by drawing the attention of and therefore attracting more talented people to improve the business's productivity. Therefore, it is important to fill gaps in the workplace environment with a passion for unity, teamwork and collective productivity to benefit the organisation.

1.3. Business Case for Diversity Prediction

In the 1990s, the discussion on diversity shifted to highlight its supporting effects on businesses and the diverse workforce community. This has led to a recognition that diversity in business is positive and in fact has been influential in generating young business tycoons and genius leaders (Jackson et al., 1991). The positive effects of diversity in the business world have been evaluated through corporate literature, where it has been demonstrated with the help of social procedures and protocols. For example, desirability and attraction in similarity among diverse cultures make it promising to measure the performance and productivity of business operations and services. Many studies have been conducted on large organisations that are highly reputed for their committed services and management of their business with the support of a diverse workforce (Harrison et al., 2002). We can take an example from the business report of the former CEO of the multinational company "Hewlett Packard". He found an opening to prove to his business fellows, managers, and executives that management of diversity in business is essential for the efficiency of business and the benefits of labor. He further said that a diverse workforce produces improved business outputs. He also thought that the rate of business progress thus eventually develops a more competent workforce, and the organisation can integrate fully across the superior levels of the business pyramid. Diversity in business is associated with high turnover among top organisations and top business leaders (Jackson et al., 1991). Therefore, business diversity also requires special attention to be able to filter down as an important topic among policymakers, general society, research scholars, social media, and business leaders. Due to a lack of appropriate framework on the benefits of diversity in business models, it is not easy to promote the value of diversity in business activities (Kochan et al., 2003).

1.4. Benefits of a Data Driven Approach in Diversity

In the past 25 years, we can observe that diverse groups of people have immigrated to the United States of America. As a result, many small and large-scale businesses have realized they should hire diverse talent from different countries to grow their organisations and to develop their product services. Previous research studies have shown that companies with a diverse workforce out-performed outclassed and out-innovated competitors with increased margins of annual revenues (Cross & Braswell, 2017). Business performance radically improves with diversity, which provides competitive benefits to organisations in building a good bond with their customers (Moieni et al., 2022). There is, however, always a major restriction for companies to integrate foundational strategies, for example: mutuality, inclusion, and diversity, into the evaluation of benefits against cost of the organisations. It has been reported that most companies spend approximately 8 billion dollars on diversity training (Madera, 2013). However, those companies still lack methods of assessing the effectiveness of and performance of diversity strategies within their organisation (Hunt et al., 2018).

Another group of researchers, (Kochan et al., 2003) reported that when large and well-known organisations do not implement systematic analysis on the impact of diversity and diversity initiatives, it will create a considerable gap in the knowledge of their diversity within and how it impacts performance. Another researcher (Moieni & Mousaferiadis, 2022), proposed a new model in 2022 that evaluates cultural diversity across four categories. The researcher measures cultural diversity into different groups such as diverse beliefs, diverse languages, diverse ethnicity, and diverse group of other nationalities in the specific community (Moieni et al., 2022).

Certain factors are very important to stimulate data-driven approaches, such as developing diversity relative metrics to support data-driven approaches (Jayne & Dipboye, 2004). Therefore, finding and identifying diverse resources is key to avoiding costly yet ineffective strategies in organisations. In addition, the data-driven approach allows insightful data to provide a pathway for companies to explore open new business ventures and achieve their targets.

2. Data-Driven Forecasting Models Using Machine Learning to Predict Diversity

Modelling and predicting short time series is a very complex task that heavily depends on the type of data-set used. Of all the forecasting methods, the model chosen will vary based on what type and how many variables are required to be predicted, as well as the amount of noise, and randomness in the data. It therefore follows that as the number of parameters estimated, or as the noise increases, the requisite sample size will also increase (Hyndman & Athanasopoulos, 2018).

3. Applied Methods

1) Grey model GM (1, 1): Grey model 1 is considered one of the newest and most innovative models for limited data. This model is very useful because can provide accurate estimations using only 4 observations. The model is made by first applying an accumulating generation operator (AGO) to the data and then the governing differential equation of the model is solved to obtain the predicted value of the system. Lastly, the predicted value of the original data is obtained by using the inverse accumulating generation operator (IAGO). The implemented results for GM (1, 1) have been demonstrated in **Table 1**.

2) ARIMA: The ARIMA model contains three other models including autoregressive (AR), moving average (MA), and seasonal auto-regressive integrated moving average (SARIMA) (Benvenuto et al., 2020). ARIMA models are based on the notation ARIMA (p, d, q) (P, D, Q) where the second group of letters are

Table 1. Prediction for top 10 countries by GM (1, 1) model.

Year \ Countries	India	China	UK	New Zealand	Philippines	S. Africa	Viet Nam	Malaysia	Indonesia	Australia
2022	111,348	75,965	16,178	17,762	11,128	15,712	1986	6139	4193	−18,841
2023	119,386	87,507	16,686	17,710	11,474	16,622	1904	6293	4341	−19,246
2024	131,081	96,225	16,223	17,852	11,831	17,586	1827	6450	4495	−19,660
2025	143,922	105,814	16,560	17,996	12,199	18,605	1752	6611	4654	−20,082
MAPE	30%	16%	21%	176%	11%	29%	20%	10%	40%	29%

responsible for the seasonal aspect of the data. The letters p, d and q are responsible for the order of the auto-regressive part, the degree of first differences involved and the order of the moving average part (Panagiotelis et al., 2021). The number of the parameters of a model is related to the sum of the values of p + q + d, thus p + q + d + 1 observations are required to estimate a model. The results are demonstrated in Table 2.

In the study, these parameters were tuned using the auto-correlation function (ACF) graph and partial auto-correlation (PACF) correlogram, in addition to a tuning algorithm by comparing the different combinations of parameters based on MAPE.

3) Model performance criteria MAPE: The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of the prediction accuracy of a forecasting method in statistics. It usually expresses the accuracy as a ratio defined by the formula:

$$\mathrm{MAPE} = \frac{100\%}{n} * \sum_{t=1}^{n} \frac{|A_t - F_t|}{A_t} \tag{1}$$

where A_t is the actual value and F_t is the forecast value. Their difference is divided by the actual value A_t. The absolute value of this ratio is summed for every forecasted point in time and divided by the number of fitted points n.

4. Results and Discussion

We have applied two different methods of GM (1, 1) (Table 1) and ARIMA model (Table 2) to NOM (Net Overseas Migration) data to predict the NOM of 10 countries with highest migration numbers including Australia. Table 4 shows the summary statistics of NOM data. Since 2020 and 2021 are part of the pandemic, and regulations have changed drastically regarding immigration, these two years are removed from data and considered as anomalies for consistency and we have predicted NOM from 2022 to 2025. Census top 10 countries of birth prediction by leveraging NOM predicted data by ARIMA model is demonstrated in Table 3.

The statistical summary of the NOM dataset is presented in Table 4. The standard deviation is one of the critical points of the table. It is widely used to measure variability. A higher standard deviation means that data are further

Table 2. Prediction for top 10 countries by ARIMA model.

Year \ Countries	India	China	UK	New Zealand	Philippines	S. Africa	Vietnam	Malaysia	Indonesia	Australia
2022	85,014	9418	17,404	12,206	20,625	5276	8086	2049	4164	−8640
2023	86,931	18,583	17,611	13,850	21,291	5161	8072	3011	4311	−9738
2024	88,849	30,785	17,666	14,854	21,962	5052	8101	3523	4392	−11,526
2025	90,766	40,985	17,681	15,404	22,633	4958	8112	3795	4459	−13,359
MAPE (%)	16	22	26	160	12	19	11	39	19	13

Table 3. Census top 10 country of birth prediction by leveraging NOM predicted data by ARIMA model.

Year \ Countries	India	China	UK	New Zealand	Philippines	S. Africa	Vietnam	Malaysia	Indonesia	Australia
2025	1,061,940	695,401	1,022,052	NA	397,131	222,377	300,541	29,598	NA	NA
MAPE	16%	22%	27%	NA	11%	19%	11%	38%	NA	NA

away from the mean value, and the values are spread out over a broader range. In other words, data values become more dissimilar, which means frequently occurring extreme values. Thus, it is clear from the standard deviation results that there is a wider spread of the population prediction.

We have implemented both ARIMA and GM models and calculated the models' performance (MAPE), it is realized that the ARIMA model has a lower average MAPE compared to the GM model for all countries (33.7 % and 38.2%, respectively). Therefore, we use the predicted data based on ARIMA model to calculate the 2025 Census data.

Table 2 shows the results of ARIMA model prediction for ten studied countries of the research. Based on the built model from the training dataset, we predict the migration population from each country from 2022 to 2025. Based on MAPE, we can see that the model has an acceptable error for all countries, excluding New Zealand. We have excluded New Zealand later in this study as it follows another pattern. This will be addressed later in this paper.

Moreover, China and India have the highest immigration growth among all nations and South Africa is the only country that the number of immigrants slightly decreases during the period and UK has a minor increase during the same period. We have predicated the country of birth of 7 out of 10 countries in 2025 by leveraging NOM predicted data II and census data of country of birth data in 2021. The idea was that the census data in 2021 shows the country of birth of current population in Australia and NOM data shows the number of migrants from top 10 countries that will be added to this population until 2025, hence by accumulating these population per country and adding this to the population of census data in 2021, we have obtained the predicted population of migrants from 7 countries in Australia. The average prediction error rate (MAPE) for each country is the same as ARIMA's model and the highest population belongs to India and UK in 2025, while the lowest of this table belongs to Malaysia.

Three countries of New Zealand, Indonesia and Australia were not included in this prediction as there is no census data for Indonesian population in Australia in 2021, We will explain later in this paper that why New Zealand has no predictable pattern. Moreover, NOM data is not suitable for predicting the number of populations born in Australia.

In this study, those whose country of birth is New Zealand are not included here as they do not follow the same migration pattern as other countries. Under various arrangements since the 1920s, there has been a free flow of people between Australia and New Zealand. Under the Trans-Tasman Travel Arrangement introduced in 1973, Australian and New Zealand citizens are able to enter each other's country to visit, live and work indefinitely, without the need to apply for prior authority. New Zealand is the only country in the world that has such an arrangement with Australia. There are no caps on the numbers of New Zealanders who may enter under the arrangement, and the only limitations on entry relate to health and character requirements (Parliament of Australia, 2022). As of 30 June 2018, there were an estimated 568,000 New Zealand-born people living in Australia, forming 2.3 per cent of the population and representing our fourth-largest migrant community (Parliament of Australia, 2022).

There are other parameters that impact a more accurate prediction, such as government total migration cap or government policies, global situations such as pandemic, regional conflicts such as wars, global economy situation and recessions. We considered these features the same for all studied countries.

Religions and Languages

When it comes to predicting the religions and languages of Australians, available data over time is not sufficient to be used for a more accurate data prediction. Mainly because:

- People's beliefs and also the languages they mainly speak may vary in time (Unlike country of birth which is always constant data for each person) Unlike their country of birth, which is immutable and does not change, language and worldviews and beliefs may be modified over the course of a person's life—languages can be learned, and beliefs may change. Therefore, further data is required to predict this behaviour. The statistical summary of the prediction of the top 6 languages in Australia is presented in **Table 4**.
- The languages and religions of migrants to Australia are hard to predict. While NOM data exists, and we know the language and religious demographics of each country, we cannot use this data to know the exact details of people entering Australia
- When it comes to languages, in the census, only the data for the language spoken at home is asked. While this can be considered a good indicator of spoken languages, people may speak more than one language at home or at work at different proficiency levels. This needs to be considered for a more accurate data analysis. Prediction for top 6 languages by leveraging Census data, NOM and country of birth data has been demonstrated in **Table 5**.
- Same as language, people may feel aligned to more than one religion (such as when they are raised by parents of different faiths). The prediction of top 4 beliefs in Australia has been presented in **Table 6**. It shows steady growth in the number of people who consider No religion as their main worldview
- The datasets for both language and religion are not inclusive enough to consider the true linguistic or religious diversity of Australians. The disparity or distance between different languages or religions needs to be considered as well.

Table 4. Summary statistics of datasets

	Mean	Std. Dev.	Min.	Median	Max
India	15,923.72	19,740.17	844	73,190	73,190
China	14,008.6	14,415.74	1167	5086	50,020
UK	17,705	8780.73	7440	14,587	36,994
New Zealand	14,105.1	8171.17	930	13,232.5	34,540
Philippines	7040.3	4464.05	2013	5618.5	18,400
South Africa	4546.62	4546.62	1021	4180.5	12,870
Vietnam	6533.25	3244.02	1502	6984	13,248
Malaysia	4111.57	2887.23	931	3445	12,180
Indonesia	1932.92	1105.09	83	1565	4221
Australia	−13126.5	8960.22	−46,989	−10,013.5	−2280

Table 5. Prediction for top 6 languages by leveraging Census data, NOM and country of birth data.

Year \ Languages	English	Mandarin	Arabic	Vietnamese	Cantonese	Punjabi
2021	18,304,407	686,415	355,920	3,330,497	305,074	228,805
2022	18,662,400	686,469	617,193	339,421	312,480	231,859
2023	18,885,600	686,469	620,318	339,421	312,517	231,859
2024	19,130,400	686,469	623,745	339,422	312,558	348,859
2025	19,368,000	686,469	627,072	339,421	312,598	348,859

Table 6. Prediction for top 4 worldviews by leveraging Census data, NOM and country of birth data.

Year \ Languages	Christianity	No religion	Not stated	Islam
2021	12,135,786	9,894,144	1,777,854	824,512
2022	12,208,320	9,953,280	2,144,400	1,159,937
2023	2,354,330	10,758,735	2,165,790	1,169,857
2024	12,514,470	10,889,295	2,189,250	1,180,737
2025	12,669,900	11,016,015	2,212,020	1,191,297

Based on Australian Bureau of Statistics (ABS) Population Projections, Main Projection Series are total fertility rate, Net overseas migration, and Life expectancy at birth (Parliament of Australia, 2022). Based on (Parliament of Australia, 2022), Australia's population will be 25.92 million by 2022, 26.23 million by 2023, 26.57 million by 2024 and 26.9 million by 2025. These numbers used to estimate (not predict) the distribution of languages and religions across Australia from 2022 to 2025 and that is with the assumption that same percentage of people will follow those religions and keep communicating with the same languages. Once there is more data on ethnolinguistics and religions of people, a more accurate data prediction on this area can be calculated. In **Table 5** and **Table 6**, we have calculated the language and worldviews based on mentioned assumption.

When considering diversity prediction of a community using available methods, some technical and non-technical points may impact the study:

- The dynamic nature of human beings. People may change their cultural attributes over time, such as changing their beliefs or learning a new language.
- The process of collecting diversity data is a fairly new concept and its data structure has changed many times since collection has started.
- The available datasets on the diversity of communities, are not inclusive and comprehensive and the disparity and distance between elements are not usually addressed.
- The diversity data collection on national scale, usually happens every few years meaning the changes during that period will be missed in the collection of data.

Some organisations do this on an annual basis, which still increases the error rate while doing diversity forecasting using ARIMA or Grey methods.

5. Future Work

In this research project, two machine learning models were used on the training data, GM (1, 1) suitable for small data and ARIMA model which is a strong model which had a higher accuracy for the purpose of this research. GM (1, 1) is not suitable for those data with high deviation, so non-linear Grey Model such as GM (1, 1) may have a better result and needs to be considered in future work. We used NOM (Net Overseas Migration) data to train our data forecasting model based on small data and then applied that on census data from Australia. We then predicted the future diversity of Australians by 2025 including country of birth, language and religions.

As data around religions and languages of residents of Australia was not sufficient, more comprehensive data is needed for a more accurate calculation. In future work, we need to gather data on people who change their religions and more data around those who learn a new language.

More data is also required to calculate the NOM data for Australians as NOM data is based on country of birth, but there is no such data available for those Australians who leave Australia.

In future work, ideally we would aim to broaden the concept of diversity prediction by applying it to other demographic fields like disability, gender or gender gap, level of education, age, among others, rather than limiting to the data available in the NOM data-set. Also, we would apply this model to predict the gap between diversity of a community and the community around it (also known as mutuality). The algorithm to be developed in a tool named Diversity Atlas.

Acknowledgements

This research was supported by Cultural Infusion PTY Ltd. We thank Ras Amirzadeh, Michael Walmsley, Quincy Hall, Duncan Foster, Jane Felstead and Nicola Diomides who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this article. We also thank Jane Felstead for editorial support and advice.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Baldwin, K., & Huber, J. D. (2010). Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision. *American Political Science Review, 104,* 644-662. https://doi.org/10.1017/S0003055410000419

Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the Arima Model on the Covid-2019 Epidemic Dataset. *Data in Brief, 29,* Article ID: 105340. https://doi.org/10.1016/j.dib.2020.105340

Brubaker, R. (2013). Language, Religion and the Politics of Difference. *Nations and Nationalism, 19,* 1-20. https://doi.org/10.1111/j.1469-8129.2012.00562.x

Cletus, H. E., Mahmood, N. A., Umar, A., & Ibrahim, A. D. (2018). Prospects and Challenges of Workplace Diversity in Modern Day Organizations: A Critical Review. *HOLISTICA-Journal of Business and Public Administration, 9,* 35-52. https://doi.org/10.2478/hjbpa-2018-0011

Cox, T. H., Lobel, S. A., & McLeod, P. L. (1991). Effects of Ethnic Group Cultural Differences on Cooperative and Competitive Behavior on a Group Task. *Academy of Management Journal, 34,* 827-847. https://doi.org/10.2307/256391

Cross, S. & Braswell, P. (2017). A Data-Driven Approach to Hiring More Diverse Talent. *Harvard Business Review, 10.*

Cryer, J., & Chan, K. (2008). *Time Series Analysis*. Springer. https://doi.org/10.1007/978-0-387-75959-3

Green, K. A., Lopez, M., Wysocki, A., & Kepner, K. (2022). Diversity in the Workplace: Benefits, Challenges, and the Required Managerial Tools. Hr022/hr022, 7/2002, EDIS.

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, Teams, and Task Performance: Changing Effects of Surface- and Deep-Level Diversity on Group Functioning. *Academy of Management Journal, 45,* 1029-1045. https://doi.org/10.2307/3069328

Helpas, N. (2022). *Cultural Diversity and National Performance*. WP5/06 Search Working Paper.

Hunt, V., Prince, S., Dixon-Fyle, S., & Yee, L. (2018). *Delivering through Diversity*. McKinsey & Company.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.

Jackson, S. E., Brett, J. F., Sessa, V. I., Cooper, D. M., Julin, J. A., & Peyronnin, K. (1991). Some Differences Make a Difference: Individual Dissimilarity and Group Heterogeneity as Correlates of Recruitment, Promotions, and Turnover. *Journal of Applied Psy-*

chology, 76, 675-689. https://doi.org/10.1037/0021-9010.76.5.675

Jayne, M. E., & Dipboye, R. L. (2004). Leveraging Diversity to Improve Business Performance: Research Findings and Recommendations for Organizations. *Human Resource Management, 43,* 409-424. https://doi.org/10.1002/hrm.20033

Jonsen, K., Maznevski, M. L., & Schneider, S. C. (2011). Special Review Article: Diversity and Its Not So Diverse Literature: An International Perspective. *International Journal of Cross Cultural Management, 11,* 35-62. https://doi.org/10.1177/1470595811398798

Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., Leonard, J., Levine, D., & Thomas, D. (2003). The Effects of Diversity on Business Performance: Report of the Diversity Research Network. *Human Resource Management, 42,* 3-21. https://doi.org/10.1002/hrm.10061

Krech, V., Hero, M., Huber, S., Ketola, K., & Traunmuller, R. (2013). Religious Diversity and Religious Vitality: New Measuring Strategies and Empirical Evidence. *Interdisciplinary Journal of Research on Religion, 9,* Article 3.

Lorenzo, R., Voigt, N., Schetelig, K., Zawadzki, A., Welpe, I., & Brosi, P. (2017). *The Mix That Matters: Innovation through Diversity.* The Boston Consulting Group.

Madera, J. M. (2013). Best Practices in Diversity Management in Customer Service Organizations: An Investigation of Top Companies Cited by Diversity Inc. *Cornell Hospitality Quarterly, 54,* 124-135. https://doi.org/10.1177/1938965513475526

Moieni, R., & Mousaferiadis, P. (2022). Analysis of Cultural Diversity Concept in Different Countries Using Fractal Analysis. *The International Journal of Organizational Diversity, 22,* 43-62. https://doi.org/10.18848/2328-6261/CGP/v22i01/43-62

Moieni, R., Mousaferiadis, P., & Pateel, P. (2022). An Analytical Approach to Measure the Cultural Diversity Mutuality between Two Communities. *NeuroQuantology, 20,* 105-120.

Page, S. E. (2007). Making the Difference: Applying a Logic of Diversity. *Academy of Management Perspectives, 21,* 6-20. https://doi.org/10.5465/amp.2007.27895335

Panagiotelis, A., Athanasopoulos, G., Gamakumara, P., & Hyndman, R. J. (2021). Forecast Reconciliation: A Geometric View with New Insights on Bias Correction. *International Journal of Forecasting, 37,* 343-359. https://doi.org/10.1016/j.ijforecast.2020.06.004

Parliament of Australia (2022). https://www.aph.gov.au/