Crystal Ball for the Impact of Anthropogenic Climate Change on Global Air Quality PM 2.5

Shan Shan¹

¹Department of Sociology, Zhejiang University

July 23, 2023

1 Crystal Ball for the Impact of Anthropogenic Climate Change on Global Air Quality PM 2.5: a

2 Machine Learning Method

3 Shan Shan

- 4 Department of Sociology, Zhejiang University, Hangzhou 310058, China
- 5 Corresponding author: Shan Shan (<u>shshan@zju.edu.cn</u>)

6 Key Points:

- 7 List up to three key points (at least one is required)
- Key Points summarize the main points and conclusions of the article
- Each must be 140 characters or less with no special characters or acronyms.
- 10 1. Significant shifts in PM2.5 emissions are projected in relation to complex social factors.
- In terms of model efficacy, the Random Forest and Support Vector Regression
 methodologies stood out for this scenario.
- Machine learning revealing the importance of interdisciplinary approaches in addressing
 PM2.5 pollution.
- 15
- 16

18 Abstract

19 Air pollution caused by PM2.5 particles is a major global concern, particularly for human health.

20 This study used machine learning tools to uncover the social factors influencing PM2.5 emissions.

- 21 Text mining techniques were employed to extract key variables from previous research databases
- related to the target variable PM2.5 air pollution and its determinants. Four important features were
- derived, encompassing a wide range of factors, including PM2.5 air pollution levels, population
- 24 metrics, GDP per capita, military expenditure, health expenditure, and environmental features. The 25 study identifies significant changes in PM2.5 emissions related to health expenditure and economic
- 26 contributions, emphasizing the need for interdisciplinary efforts to address this global problem.
- 27 From a technical standpoint, machine learning feature extraction was deployed to pinpoint four
- critical factors with significant influence on air quality. In terms of model efficacy, the Support
- 29 Vector Regression method stood out. This technique excelled in producing accurate predictions
- 30 and understanding the correlation between key social factors and global exposure to PM2.5.
- 31

32 Plain Language Summary

33

34 Classified as the most hazardous air pollutant globally, PM2.5 draws significant interest 35 from scholars and the public, especially in terms of human health. Leveraging machine learning 36 techniques, this study investigates the determinants of these emissions in multi-social dimensions. 37 The conclusions denote that the complexity and implications of airborne particulate pollution are 38 shaped by multifaceted social variables economically and politically. Machine learning methods 39 unravel multi-dimensional societal causes in PM2.5 emissions within the context of health 40 expenditure and economic contributions from urban population growth, agriculture, forestry, and 41 fishing, value added (% of GDP), manufactures exports (% of merchandise exports), current health 42 expenditure (% of GDP), asserting the significance of an interdisciplinary approach in combating 43 this global issue.

44

45 **1 Introduction**

46

47 Without a crystal ball, it is hard for a research to accurately predict the long-term effects 48 of climate change, since such predictions involve a degree of complexed social uncertainty. Air 49 pollution as an indicator of climate change is a pressing global issue that poses significant risks to 50 human health and the environment(OECD, Awe, 2022). Among the various air pollutants, particulate matter with a diameter of 2.5 micrometers or less (PM2.5) is classified as the most 51 52 hazardous one. Its adverse effects on respiratory and cardiovascular health have been widely documented, making it a matter of utmost concern for public health authorities and policymakers 53 54 worldwide(Cohen et al. 2017). Understanding the determinants of PM2.5 emissions and accurately 55 predicting their future patterns are crucial for developing effective strategies to mitigate air 56 pollution and protect human well-being(Karagulian, 2023:1).

57 Traditionally, the analysis of PM2.5 emissions and their associations with socioeconomic 58 factors has relied on static linear regression models. However, the complexity of airborne 59 particulate pollution and its implications extend beyond simple linear relationships. In recent years, 60 machine learning techniques have emerged as powerful tools for deciphering the intricate 61 dynamics of complex systems, including environmental phenomena. Leveraging the capabilities

62 of machine learning, this study aims to unravel the determinants of PM2.5 emissions to offer an

63 example for more policy makers.

The objective of this investigation is to go beyond the limitations of traditional static linear regression models and provide a more profound understanding of the societal causes of PM2.5 pollution. By applying machine learning methods, reserachers can capture non-linear relationships and interactions among various social variables, both economically and politically, that shape the complexity of airborne particulate pollution. This research contributes to the growing body of knowledge on air pollution by delivering more precise forecasts and shedding light on the multifaceted factors influencing PM2.5 emissions.

71 In this study, a novel and efficient method was implemented to facilitate literature reviews 72 by using text mining techniques and automated database interactions. Leveraging the Biopython 73 library and the National Center for Biotechnology Information (NCBI) Entrez database, the 74 employed text mining techniques fetch relevant articles associated with PM2.5 air pollution, climate change, and global air quality and extract critical variables from the retrieved literature, 75 76 leading to the identification of 17 key features that spanned a multitude of factors, including PM2.5 77 air pollution levels, population metrics, GDP per capita, military expenditure, health expenditure, 78 and environmental indicators.

79 Furthermore, this study highlights the significance of an interdisciplinary approach in addressing the global issue of PM2.5 pollution. By identifying the most pronounced shifts in 80 81 PM2.5 emissions within the context of health expenditure and economic contributions from sectors 82 such as urbanization, economic growth from agriculture, forestry, and fishing, manufactures 83 exports (% of merchandise exports), and current health expenditure (% of GDP). The study indicates the need for collaboration between different fields of expertise. Only through a 84 85 comprehensive understanding of the societal factors influencing PM2.5 emissions can effective 86 strategies and policies be developed to combat air pollution and safeguard human health.



Figure 1. The diagram of analysis design. This research design, firstly, an exploration
of multi-dimensional social causes was undertaken using text mining techniques,
allowing for an analysis of numerous academic works, the identification of popular
topics, and their classification into four distinct categories representing different
dimensions of social data. Secondly, machine learning was employed to quantify and
highlight the most critical features. Lastly, a comparative study of commonly used
models was conducted to compare different model fitness.

96

97 2 Materials and Methods



To investigate the influence of human-induced climate change on global PM2.5 air quality (Fig.2), a comprehensive dataset was collected from reputable sources such as the World Bank and the OECD global data. The referenced data from OECD is Air pollution exposure, which pertains to population exposure levels exceeding 10 micrograms per cubic meter, and these figures 103 are presented as yearly averages. The World Bank development indicator (WDI) encapsulates









107 Text mining techniques were employed to extract key variables from previous research 108 related to PM2.5 air pollution and its determinants. Four social-important features' categoorites 109 are derived, encompassing a wide range of factors including PM2.5 air pollution levels, population 110 metrics, GDP per capita, military expenditure, health expenditure, and environmental indicators.

111 The study utilizes the Biopython library to interact with the National Center for Biotechnology Information (NCBI) Entrez database, retrieving articles related to climate change 112 113 and global air quality. The developed code snippet capitalizes on the functionalities of the Biopython library and the NCBI Entrez database to conduct guery-based searches for articles. The 114 Entrez module from the Bio package is employed, ensuring compliance with NCBI guidelines for 115 116 accessing resources. Furthermore, the esearch function from the Entrez module retrieves search results from the 'pubmed' database in XML format. Detailed information of identified articles is 117 extracted using the fetch details function, which takes a list of article IDs as input. The efetch 118 119 function from the Entrez module is employed to fetch these details from the 'pubmed' database in XML format. The practical application of this code is demonstrated by searching for articles 120 related to "Climate Change, Global Air Quality, and Society", storing the ID list of matching 121 122 articles in the id list variable. The fetch details function then retrieves the detailed information of 123 these articles.

Through the process of text mining, pertinent features are classified into four distinct categories: environmental indicators, military and security indicators, economic indicators, healthcare indicators. This classification facilitated a more focused analysis of the socio-economic factors contributing to PM2.5 emissions. Utilizing open data resources such as the World Bank and OECD open data source, associated sub-indicators were amassed as detailed below:

- 129
- 130

131	Environmental Indicators:
122	DM2.5 sin nellution experime mean empired experime (micro enough experime
132	• FW2.5 all pollution exposure, mean annual exposure (interograms per cubic meter)
134	• CO2 emissions (kt)
154	
135	Military and Security Indicators:
136	Armed forces personnel, total
137	• Armed forces personnel (% of total labor force)
138	• Military expenditure (% of GDP)
139	Economic Indicators:
140	• Foreign direct investment, net outflows (% of GDP)
141	Urban population
142	• Urban population growth (annual %)
143	• Agriculture, forestry, and fishing, value added (% of GDP)
144	• Services, value added (% of GDP)
145	• Trade in services (% of GDP)
146	• GDP per capita (constant 2015 US\$)
147	• GDP per capita growth (annual %)
148	 Foreign direct investment, net inflows (% of GDP)
149	• Manufacturing, value added (% of GDP)
150	• Manufactures exports (% of merchandise exports)
151	• Exports of goods and services (% of GDP)
152	• Goods and services expense (% of expense)
153	Urban population (% of total population)
154	Healthcare Indicators:
155	• Life expectancy at birth, total (years)
156	• Lifetime risk of maternal death (%)
157	• Lifetime risk of maternal death (1 in: rate varies by country)
158	• Current health expenditure (% of GDP)
159	• Current health expenditure per capita (current US\$)
160	• Current health expenditure per capita, PPP (current international \$)
161	• Domestic general government health expenditure (% of GDP)
162	Considering the raw data's degree of comprehensiveness and uniformity, 19 sub-indicators
163	were retained for further analysis (refer to Table 1).
164	

Table 1Basic information for nineteen sub-inicators

Series Name	Series Code
PM2.5 air pollution exposure	ExpoPM2.5
Agriculture, forestry, and fishing, value added (% of GDP)	NV.AGR.TOTL.ZS
Military expenditure (% of GDP)	MS.MIL.XPND.GD.ZS
Services, value added (% of GDP)	NV.SRV.TOTL.ZS
Trade in services (% of GDP)	BG.GSR.NFSV.GD.ZS
CO2 emissions (kt)	EN.ATM.CO2E.KT
GDP per capita (constant 2015 US\$)	NY.GDP.PCAP.KD
Foreign direct investment, net inflows (% of GDP)	BX.KLT.DINV.WD.GD.ZS
Manufacturing, value added (% of GDP)	NV.IND.MANF.ZS
Manufactures exports (% of merchandise exports)	TX.VAL.MANF.ZS.UN
Tariff rate, applied, simple mean, manufactured products (%)	TM.TAX.MANF.SM.AR.ZS
Textiles and clothing (% of value added in manufacturing)	NV.MNF.TXTL.ZS.UN
Exports of goods and services (% of GDP)	NE.EXP.GNFS.ZS
Goods and services expense (% of expense)	GC.XPN.GSRV.ZS
Total debt service (% of GNI)	DT.TDS.DECT.GN.ZS
Urban population growth (annual %)	SP.URB.GROW
Urban population (% of total population)	SP.URB.TOTL.IN.ZS
Armed forces personnel (% of total labor force)	MS.MIL.TOTL.TF.ZS
Armed forces personnel, total	MS.MIL.TOTL.P1

167 2.2 Data splitting

The initial steps in the analysis involve splitting the dataset into training and testing subsets. The `train_test_split` function from the scikit-learn library is employed for this purpose. By randomly dividing the dataset, models are trained on a portion of the data while their performance is assessed on unseen data. The study allocates 70% of the data for training and 30% for testing. A fixed random seed of 19 ensures consistency across iterations, facilitating the development of robust models and effective assessment of their predictive capabilities (Bisong, 2019:251, 2019:289).

175 2.3 The distribution and probability property of the data

176 The distribution properties rely on certain distributional assumptions about the 177 data(Demirtas & Yucel, 2008).

This analysis aims to delve into the detailed distribution characteristics of our dataset's variables. The nature of the data is better understood by examining the distribution and quantilequantile plots for each column in the dataframe(NIST/SEMATECH, 2022). The generated plots provide insights into the distribution and deviation from normality for each column in the DataFrame. The histogram and fitted curve show the shape of the distribution, and the probability 183 plot helps evaluate how well the data aligns with a normal distribution indicates there is no 184 significant differences in the distribution of our features between our training and testing 185 sets.(Fig.3).

186



- 187 188
- 189 190

Fig.3. The distribution and probability property of the data

191 Ensuring that the training and testing sets are similarly distributed guarantees that the 192 model can effectively learn pertinent patterns during the training phase and accurately apply them 193 to the test data. (Bisong, 2019). Thus kernel density plots are utilized for the visual examination

194 of the distribution of each feature in the training and testing subsets. Kernel Density Estimation 195 (NIST/SEMATECH, 2022) also works as a method that assists in smoothing a histogram and 196 facilitates data visualization via a continuous probability density curve across one or more 197 dimensions. The outcomes from this comparative examination shows that the test data mirrors the 198 training data properly. Through the analysis of these scatter plots and histograms, the aim is to 199 identify any intriguing relationships or patterns in the data. There is no significant skewness or 200 abnormalities in the feature distribution and no additional data cleaning steps or motivate 201 transformations, thus further training, validation, and assessment procedures could be applied 202 (Fig.4).



- 203
- 204

Figure 4. The distribution and probability property of the data with KDE

205



The method used for the exploratory data analysis is described, with a focus on understanding the relationships among the variables in the dataset. The data analysis process involves the construction of correlation matrices to visualize the linear relationships between pairs of variables to the target variable air quality. The correlation coefficient, ranging from -1 to 1, indicates negative and positive correlations, with 0 indicating no linear relationship. For the overall interrelationships among the variables, a heatmap of the correlation matrix is created for air quality "ExpoPM2.5"(Fig.5). A positive correlation between two variables is represented by a bright color,

- and a negative correlation is a darker color, which expedites the process of feature selection and
- 215 helps identify potential collinearity issues.





Fig.5. Heatmap of the overall interrelationships among the variables

- 218
- 219 2.5 Four important features

220 A correlation threshold of 0.5 is established to streamline the focus. This threshold signifies 221 the minimum absolute correlation a variable must exhibit with the `ExpoPM2.5` to be considered 222 significant for the analysis. Variables that possess correlation values less than this threshold with 223 `ExpoPM2.5` are regarded as less significant and are therefore removed from the dataset. This 224 elimination process simplifies subsequent analysis and model development. A threshold of 0.5 225 suggests an interest in variables that possess a moderate to strong positive or negative connection 226 with 'ExpoPM2.5'. This interpretation of correlation strength is based on a common guideline, 227 although the threshold choice should ideally be influenced by the specifics of the research context 228 and the nature of the data. After exclusion, four crucial features remain - 'Urban population growth 229 (annual %)', 'Agriculture, forestry, and fishing, value added (% of GDP)', 'Manufactures exports (% of merchandise exports)', and 'Current health expenditure (% of GDP)'. These features, which 230 231 exhibit a significant correlation with 'ExpoPM2.5', will be the central focus of subsequent analysis 232 and model construction. In addition, correlation does not equate to causation. A high correlation

- 233 between predictors might suggest multicollinearity, which could influence the performance and
- 234 interpretability of certain types of statistical models(Fig.6).



- Fig.6. Four features with significant correlations with air quality
- 237

238 2.6 Machine learning modeling and analysis

Common machine learning techniques were implemented including Ridge Regression, Decision Tree Regressor, and Random Forest Regression for the comparasion. These models were trained and evaluated using appropriate metrics to assess their performance in predicting PM2.5 levels. The scores from the models provided empirical evidence of the validity of the analysis and the effectiveness of machine learning methods in studying PM2.5 air quality.

244 2.7 Models evaluation

245 The study conducted a comparative analysis of different machine learning tools to develop 246 a predictive model for our dataset. The models were evaluated based on their R-squared scores, 247 which measure the goodness of fit to the data. Among the models tested, Ridge Regression 248 achieved the highest R-squared score of 0.989079, closely followed by Support Vector Regression 249 with a score of 0.988993. These models demonstrated excellent predictive capabilities and 250 effectively captured the relationships between the features and the target variable. The Decision 251 Tree Regressor performed moderately well with a score of 0.845267, while the Random Forest 252 Regression obtained a lower score of 0.685836. Based on these findings, Ridge Regression and 253 Support Vector Regression are recommended as strong models for predictive modeling in our

- dataset, while further refinement may be needed for the Decision Tree Regressor and RandomForest Regression models.
- Table 2. Machine learning methods performance evaluation. The score herewithin refers to the term general to the metrics used to evaluate the performance and effectiveness of the model. (James et al. 2013).

Machine Learning Methods	Score
Ridge Regression	0.989079
Support Vector Regression	0.988993
Decision Tree Regressor	0.845267
Random Forest Regression	0.685836

The suggestion for opting for one or more models, based on the performance evaluation provided, is contingent upon a multitude of aspects, which encompass the particular objectives, requisites, and limitations associated with the task in question. Here are several points to weigh when advocating for a machine learning model in this study¹ (Table 3) :

263 Table 3. The general comparison of four machine learning models used in this study

Option description	Mechanics	Prons and cons
Support Vector Regression (SVR)	Employs kernel functions to transform data into higher dimensions, and then it tries to minimize the error within a defined margin. It is an advanced regression algorithm that aims to find an optimal hyperplane to fit the data points in a high-dimensional space. SVR is particularly effective for datasets with complex, non-linear relationships.	Pros:High predictive accuracy; Effective for complex, non-linear relationships. Cons: Computationally intensive, especially for large datasets; Model interpretability is limited.
Ridge Regression	Minimizes the sum of squared residuals with an added penalty proportional to the square of the magnitude of the coefficients.	Pros: Handles multicollinearity well; Simpler and computationally efficient. Cons: Might not perform well with non-linear data; Model complexity can be increased due to the introduction of regularization.
Decision Tree Regression	Splits the dataset into subsets based on feature values, and this process is recursively repeated until the tree reaches a predefined depth or purity.	Pros: High interpretability; Can capture non-linear relationships. Cons: Prone to overfitting, especially with complex datasets; Can create overly complex trees.
Random Forest Regression	Creates a set of decision trees from randomly selected subsets of the training set and averages their predictions.	Pros: High predictive accuracy; Less prone to overfitting compared to a single decision tree. Cons: computationally more intensive than a single decision tree. Interpretability is less compared to a single decision tree but better than SVR.

264

3. Results

266 3.1 Text mining: uncovering social-environment trends in NCBI Entrez database outputs

267 Understanding and combating PM2.5 pollution requires a comprehensive understanding of268 its societal causes and interdisciplinary collaboration.

In the present scientific investigation, three major contributions were yielded. The first contribution comprised a comprehensive examination of multi-dimensional social causes utilizing text mining methodologies. This comprehensive examination facilitated a rigorous analysis of a substantial volume of academic literature, leading to the identification of prevailing themes, which were then systematically classified into four unique categories, each representing a separate dimension of social data.

The study utilizes the Biopython library and the NCBI Entrez database, provides an example for literature retrieval and analysis. The execution of this code snippet allows for efficient retrieval of article details, including titles, from the 'pubmed' database, facilitating in-depth analysis and synthesis of research findings. It offers a tool for researchers investigating the complex relationship between climate change and global air quality, enabling the retrieval and analysis of a large volume of articles.

281 3.2 Featuring analysis: mapping the dimensions of air quality research outputs.

Stemmed from the utilization of machine learning techniques which were instrumental in quantifying and underlining the most salient features of the social data under investigation. The results of the analysis revealed the significant impact of socio-economic factors on global PM2.5 air quality. Through machine learning modeling and feature engineering, the study identified the four most influential features contributing to PM2.5 emissions: urbanization, agriculture, forestry, and fishing, value added (% of GDP); Manufactures exports (% of merchandise exports); and Current health expenditure (% of GDP).

289 3.3 Modeling evaluation: Ridge Regression and Super Vector Regression

290 In this research, Ridge Regression was employed as a machine learning method to analyze 291 the dataset. The Ridge Regression model is a variant of linear regression that includes a 292 regularization term to control the complexity of the model. By using the scikit-learn library, a GridSearchCV was performed to find the best value for the regularization parameter, a. The 293 parameter grid consisted of different a values, and the model was evaluated using a cross-294 295 validation strategy with 3 folds. The best parameter value was determined as the one that yielded 296 the highest coefficient of determination (R2) score. The results showed that the best parameter for 297 the Ridge Regression model was α =0.01, with a corresponding R2 score of 0.987. This indicates 298 that the model with the selected α value achieved a high level of prediction accuracy and 299 effectively captured the underlying patterns in the data.

300 The Support Vector Regression (SVR) model was trained using various kernel functions, including linear, polynomial, radial basis function (RBF), and sigmoid. These different kernel 301 302 functions allow the SVR model to capture different types of non-linear relationships in the data. 303 The model was also tuned using GridSearchCV, which systematically explores different combinations of hyperparameters, such as the regularization parameter (C) and the kernel 304 305 coefficient (gamma), to find the best configuration for the SVR model. The performance of the 306 model was evaluated using the R2 score, which measures the proportion of variance in the target variable that is explained by the model. Through hyperparameter optimization and R2 score 307

¹ The recommendation is not solely based on the score but also on the specific use case and requirements. Often, it is advisable to try multiple models and perform cross-validation to see how they perform on unseen data before making a final decision. Additionally, considering the interpretability, training time, and complexity in addition to accuracy is important in making an informed choice (Linardatos, Papastefanopoulos, & Kotsiantis, 2020).

308 evaluation, the optimal combination of kernel function, regularization parameter, and kernel 309 coefficient for the SVR model can be determined, resulting in a precise and robust predictive model.

310 **4. Conclusion**

The highlights the pressing global issue of PM2.5 air pollution and its detrimental effects on human health and the environment. Traditional linear regression models have limitations in capturing the complexity of airborne particulate pollution. To overcome these limitations, machine learning techniques were employed to analyze the dataset and uncover the determinants of PM2.5 emissions.

Through text mining techniques and automated database interactions, multi-dimentional variables were extracted from previous research, resulting in the identification of 18 key features encompassing four socio-economic flevels, economicaly and politically.

Machine learning methods, such as Ridge Regression, Decision Tree Regressor, and Random Forest Regression, were utilized to predict PM2.5 contributors' significance and assess the performance of the models. The results demonstrated the efficacy of Super Vector Regression and Ridge Regression's model fitness in understanding and forecasting PM2.5 air pollution.

The study emphasizes the need for an interdisciplinary approach to address the global issue of PM2.5 pollution. The identified socio-economic factors, such as health expenditure and economic contributions, provide valuable insights for policymakers and researchers in developing effective strategies and policies to mitigate air pollution and protect human health.

327 Overall, this research contributes to the understanding of the complex relationships 328 between socio-economic variables and PM2.5 emissions. By employing machine learning 329 techniques and considering multiple dimensions, policymakers can make informed decisions and 330 implement interventions to combat PM2.5 air pollution on a global scale.

331 332

333 Reference

334 335

- 336 Awe, Y. (2022). World Bank Climate Explainer Series. Retrieved from
- https://www.worldbank.org/en/news/feature/2022/09/01/what-you-need-to-know-about-climate change-and-air-pollution
- 339

340 Bisong, E., & Bisong, E. (2019). Introduction to Scikit-learn. Building Machine Learning and

- 341 Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners, 215 342 229.
- 343
- Bisong, E., & Bisong, E. (2019). More supervised machine learning techniques with scikit-learn.
- 345 Building Machine Learning and Deep Learning Models on Google Cloud Platform: A
- Comprehensive Guide for Beginners, 287-308.
- 347

348 Bisong, E., & Bisong, E. (2019). Regularization for Linear Models. Building Machine Learning

and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners,251-254.

- 352 Bolleyer, N., & B.rzel, T. A. (2010). Non-hierarchical policy coordination in multilevel systems. 353 Eur. Polit. Sci. Rev., 2, 157–185. 354 355 Breiman, L. (1996). Bagging predictors. Machine Learning, 24, 123–140. 356 https://doi.org/10.1007/BF00058655. 357 358 Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32. Available: 359 https://doi.org/10.1023/A:1010933404324. 360 361 Bohr, Jeremiah, & Dunlap, Riley E. (2018). "Key Topics in Environmental Sociology, 1990-362 2014: Results from a Computational Text Analysis." Environmental Sociology, 4(2), 181–195. 363 364 Boswell, Terry. (1989). "Colonial Empires and the Capitalist World-Economy: A Time Series 365 Analysis of Colonization, 1640–1960." American Sociological Review, 54(2), 180–196. 366 367 Boswell, Terry, & Dixon, William. (1990). "Dependency and Rebellion: A Cross-National 368 Analysis." American Sociological Review, 55(4), 540-559. 369 370 Bradford, John Hamilton, & Stoner, Alexander M. (2017). "The Treadmill of Destruction in 371 Comparative Perspective: A Panel Study of Military Spending and Carbon Emissions, 1960-372 2014." Journal of World-Systems Research, 23(2), 298-325. 373 374 Brady, David, Beckfield, Jason, & Seeleib-Kaiser, Martin. (2005). "Economic Globalization and 375 the Welfare State in Affluent Democracies, 1975–2001." American Sociological Review, 70(6), 376 921-948. 377 378 Brady, David, Beckfield, Jason, & Zhao, Wei. (2007). "The Consequences of Economic 379 Globalization for Affluent Democracies." Annual Review of Sociology, 33, 313–334. 380 381 Braswell, Taylor. (2022). "Extended Spaces of Environmental Injustice: Hydrocarbon Pipelines 382 in the Age of Planetary Urbanization." Social Forces, 100(3), 1025–1052. 383 384 Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., ... & Forouzanfar, 385 M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to 386 ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The 387 lancet, 389(10082), 1907-1918. 388 389 Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality 390 assumption when multiply imputing non-Gaussian continuous outcomes: a simulation 391 assessment. Journal of Statistical Computation and Simulation, 78(1), 69-84. 392 393 Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. 394 395 Hill, Terrence D., Andrew Jorgenson, Peter Ore, Kelly Balistreri, and Brett Clark. 2019. "Air 396 Quality and Life Expectancy in the United States: An Analysis of the Moderating Effect of
 - 397 Income Inequality." SSM Population Health 7:100346. doi:10.1016/j. ssmph.2018.100346.

398	
399	Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data
400	Mining, Inference, and Prediction. Springer Science & Business Media.
401	
402	
403	IOAir. (2022). World Air Quality Report. Retrieved from https://www.igair.com/us/world-most-
404	polluted-cities
405	1
406	Karagulian, F., De Vito, S., Karatzas, K., Bartonova, A., & Fattoruso, G. (2023), New
407	Challenges in Air Quality Measurements. In Air Quality Networks: Data Analysis, Calibration &
408	Data Fusion (Environmental Informatics and Modeling). Springer.
409	2 am 1 aston (2m 1 annound 1 monuments and 1/10 a ann 8). ~Frin 8 an
410	Kumar, P., Druckman, A., Gallagher, J., Gatersleben, B., Allison, S., Eisenman, T. S., Hoang,
411	U Hama S Tiwari A Sharma A Abhijith K V Adlakha D McNabola A Astell-Burt
412	T Feng X Skeldon A C de Lusignan S & Morawska L
413	1., 1 chg, M., Skoldon, M. C., de Eusighan, S., & Mordwiska, E.
414	(2019) The nexus between air pollution green infrastructure and human health Environment
415	International 133(Part A) 105181 https://doi.org/10.1016/j.envint.2019.105181
416	international, 195(1 att 14), 105101. https://doi.org/10.1010/j.envint.2019.105101
417	Lim W H Yamazaki D Koirala S Hirabayashi Y Kanae S Dadson S L & Sun F
418	(2018) Long-term changes in global socioeconomic benefits of flood defenses and residual risk
410	hased on CMIP5 climate models. Farth's Future 6(7) 938-954
420	based on example chinate models. Earth 51 dtale, 0(7), 750 754.
421	Murphy K P (2012) Machine Learning: A Probabilistic Perspective MIT Press
421	Mulphy, R. I. (2012). Machine Leanning. A Hobaomster reispective. Mit Hess.
422	Normann H F (2017) Policy networks in energy transitions: The cases of carbon capture and
423	storage and offshore wind in Norway Technol Forecast Soc Change 118 80–93
+2+ ∕125	storage and orishole which in Norway. Teenhol. Torecast. Soc. Chang., 110, 00–95.
ч <i>25</i> Л26	NIST/SEMATECH e-Handbook of Statistical Methods (2022)
420	http://www.itl.nist.gov/div898/handbook/_date
427 128	http://www.hth.inst.gov/drv696/hahdbook/, date.
420 420	Soomai S. S. MacDonald B. H. & Wells P. G. (2013) Communicating environmental
429	information to the stakeholders in coastal and marine policy making: Case studies from Nova
430	Soction and the Gulf of Maine/Pay of Fundy region Mar. Policy 40, 176, 186
431	Scotta and the Outi of Manie/Bay of Fundy region. Mar. Foncy, 40, 170–180.
432	Stolz A. & Henn M. (2015). Towards Crawling the Web for Structured Data: Pitfalls of
433	Common Crowl for E Commorco. In COLD
434	Common Crawl for E-Commerce. In COLD.
433	Tach T. T. & Dong 7 (2022) Degraggion In Artificial Intelligence with Dython (nr. 162-181)
+30 137	Singanore: Springer Singapore
+J/ /20	Singapore. Springer Singapore.
+30 120	
439 440	
440	