Topic Modeling of the Pakistani Economy in English Newspapers via Latent Dirichlet Allocation (LDA)

Fasih Ahmed, Muhammad Nawaz, and Aisha Jadoon

Abstract
This research paper explores aspects of the Pakistani economy using the Latent Dirichlet Allocation (LDA) technique. The data based on 3,000 articles were collected from two Pakistani English newspapers, Dawn and The News, (2015–2020), through Lexis Nexis database. The headlines of the news articles relevant to Pakistan’s economy, were taken into account. By employing the data-driven approach of the grounded theory, it is found that changes in policies, security preference, textile industry, the shift of energy, inflation, growth and investment, mega projects, sustainable democracy and poverty control need to be focused to overcome the challenges of Pakistan’s economy. It also reveals that mega projects like the China Pakistan Economic Corridor (CPEC) are called to boost Pakistan’s economy. The results show that smooth trading would help reduce poverty in the country.

Keywords
topic modeling, natural language processing, grounded theory, data-driven approach

Introduction
Pakistani economy has faced multiple challenges in the past. According to a World Bank report (Easterly, 2001), Pakistan systematically underperformed on social and political indicators from 1950 to 1999. These indicators include education, health, sanitation, fertility, gender equality, corruption, political instability, and democracy. Easterly terms this economic state as “growth without development.” Similarly, Hussain et al. (2018) state that the Pakistani economy’s decline started in the early 90s. Hussain (2000) writes that the ultimate determinant of a country’s economic success or failure is governance, not foreign aid. He further adds that Pakistan’s overall economic situation refers to the reality that economic development during the past 50 years has predominantly benefited a small class of elite while the majority of the population remains illiterate, impoverished, and backward. Similarly, Rehman et al. (2015) refer to literacy as an indicator of better economic growth, where Pakistan is ranked 113 among 120 literate countries. For the last decade, the economic condition of multiple sectors in Pakistan has been struggling to overcome the menace of poverty and unemployment (Gul et al., 2011; Masakure et al., 2011). The economic condition is the primary indicator for developing a country, whether it be growth, education, and employment (Hongming et al., 2020). In this context, the present study explores Pakistan’s economic dynamics (2015–2020) through a corpus-based analysis of newspapers. A corpus can be defined as a “collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis” (Tognini-Bonelli, 2001, p. 2). This approach views language as a tool that possesses the potential to discover hidden dynamics of an area by dealing with a more significant chunk of text using topic modeling (Blei et al., 2003).

Since independence to the present, many studies have explored Pakistani economy, majorly focusing on its nature, regulation, growth, challenges and policies in socioeconomic settings (Ahmed, 2017; Amjad & Burki, 2015; Hussain, 2000; Zaidi, 2005a). However, coverage of Pakistani economy in English newspapers has been nearly the neglected research area despite its significant contribution to the public knowledge regarding economic fundamentals. The study is significant for providing insights into the fundamental nature of the issues linked to Pakistan’s economy including unemployment, poverty, and the country’s growth.

1Department of Humanities, COMSATS University, Islamabad

Corresponding Author:
Fasih Ahmed, Department of Humanities, COMSATS University Islamabad, Park Road, Islamabad, 45550, Pakistan.
Email: fasih.ahmed@comsats.edu.pk

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Literature Review

Economy of a country is the backbone of its progress. Therefore, it is vital to analyze the economic challenges faced by a government and to devise a plan to overcome them (Tukhliev et al., 2020). This identification of economic challenges reveals changing trends which may result in ultimate guidance for economists and policymakers (Chandukala et al., 2008).

The economic problems of Pakistan have prolonged over almost the last four decades. The indicators reflect that in Pakistan, growth has been observed without development (Easterly, 2001). The researchers hold volatile economic policies responsible for economic challenges (Mahmood et al., 2008). The shift in policies has directly affected the other significant sectors such as energy, textile, education, tax revenues, and industrial growth (Hamid et al., 1990; Mirjat et al., 2017; White, 2015). Especially growth and development in Pakistan affected due to lack of consistent policies (Saboor et al., 2015). According to Easterly (2001), “The poor social indicators lower the productive potential of the economy and its ability to service its high debt, not to mention the loss in human welfare from having achieved so little social and political progress” (Easterly, 2001, p. 33).

A significant factor, mainly linked to Pakistan, has been the increasing number of terrorism. Terrorism has been the main challenge to the country’s growth and development (Shahbaz et al., 2013). The government struggled to overcome the menace of terrorism vice versa to a country’s economic growth (Abadie & Gardeazabal, 2008; Koh 2007). According to Hyder et al. (2015), “Besides the non-measurable loss to humans, other major economic costs of the terrorism include poverty, capital flight, destruction of infrastructure, reduction in foreign direct investment and exports, low public revenues and diversion of the development expenditure to the expenditure on law and order maintenance and so forth” (Hyder et al., 2015, p. 715). Using seven years data of World Bank, Fatima et al. (2014) dig into the detail of relationship between terrorism activities and economic growth of India and Pakistan. They concluded that terrorism impacted the overall growth in reducing the GDP (Gross Domestic Product) growth in Pakistan while India managed to overcome the impact of terrorism and its effect on GDP. Zakaria et al. (2019), analyzing the effect of terrorism on the growth of Pakistan from 1972 to 2014, highlight different variables such as foreign direct investment (FDI), domestic investment, and government spending through which terrorism influences economic growth of Pakistan. Their study concludes that in order to increase economic growth of Pakistan, more resources are to be allocated to improve law and order.

Another problem with economy is the increasing population. Peterson (2017), working on the overall economic growth of the past 200 years, concludes, “low population growth in high-income countries is likely to create social and economic problems while high population growth in low-income countries may slow their development.” This imbalance of population both in low-income may cause hurdles to the growth of economy. Being a low-income country, an increase in population negatively impacted Pakistan’s economic growth (Ahmed & Ahmad, 2016; Feeney & Alam, 2003). The data based on Pakistan economic survey from the International Financial Statistics yearbook reveal an increase of 450% increase in population from 1950 to 2001 (Afzal, 2009). It shows a negative relationship between population growth and economic advancement. On the other hand, population growth is linked to unemployment and Pakistan’s literacy rate (Ali et al., 2013). The data based on the World Bank report of 43 developing economies reveals that developing countries’ overall progress depends on population growth (Dao, 2012). This implies that population and economic development are linked to each other for the economic growth of a country. Tsen and Furuoka (2005) view population as beneficial or detrimental for the growth of economy depending on its overall situation.

Moreover, the factors like investment plans by foreign actors also affect Pakistan’s economy. However, China took the initiative for investment in the country under the umbrella of the China Pakistan Economic Corridor (CPEC) (Khan & Liu, 2019). “Economic corridors are defined as the culture of trade agreements and treaties, status, delegated legislation, and customs that govern and guide trade relations, institutions and structures, or movement of products, services and information in a geographic vicinity among people in and across borders” (Butt & Butt, 2015, p. 25). The CPEC project also attracts the regional actors to invest in Pakistan, which may, in one way or another have a positive effect on the economy of Pakistan. As “CPEC is a crucial and mutually beneficial venture that fulfills the objectives and interests of both the countries, and is also expected to enhance financial and economic cooperation between various regional actors for common development” (Butt & Butt, 2015, p. 24). Mutually beneficial adventure in the sense that China is constructing and running several special economic zones where Pakistan finds an opportunity to capitalize on the Chinese experiences involving investment, human resources and technology in establishing China-led industrial parks (Hussain & Rao, 2020). This project may help to eliminate poverty in Pakistan.

Methodology

This study investigates the economic dynamics of Pakistan. In this regard, a corpus was developed to determine the factors responsible for Pakistan’s economic conditions. The corpus was based on two Daily English newspapers, Dawn and The News. The data, the newspaper articles related to the Pakistani economy, were downloaded from the LexisNexis database in a local university of Lahore, Punjab, Pakistan. This database provided access to English Pakistani newspapers, precisely
the study 2015 to 2020. Initially, the researchers extracted 3,000 articles from the selected newspapers.

The data were extracted from LexisNexis through various searching operators. The operators like economy AND problem OR debt OR growth OR investment were employed to download the data from our targeted sources Dawn and The News (Pakistani English Newspapers) through the LexisNexis database. The operator mentioned above refers to the condition that an article for the selection must meet the terms economy and Pakistan (Weaver & Bimber, 2008). Similarly, the operator OR refers to the optional parameters for selecting the archives’ data (Lewis, 2003).

The extracted data were based on the titles of the newspaper articles. The rationale for selecting these titles of the newspapers is to present the crux of the content written in the article (Jiang et al., 2019). These headlines though provided limited data yet encompassed the complete picture of the selected articles. This data may highlight the factors responsible for the Pakistani economy.

Data Filtration

First, the data were preprocessed to filter out the specific text. This process enables the data to get rid of paralinguistic features (Denny & Spirling, 2017). These paralinguistic features do disruption to the data during the extraction process. Thus, filtration data purifies the text from the hyperlink and other unwanted features, which ultimately cause a problem during the analysis stages (Wang et al., 2014). So the file headers, footers, markup, and metadata were removed, including some additional tasks like the digits to deal with only the textual features. These features are termed as noise in natural language processing (Alasadi & Bhaya, 2017). These features, which caused noise in the text, were removed using regular expressions.

Next, the word boundaries of the data were determined by applying the tokenization process at the word level. The word-level tokenization helps in later stages to determine the text’s lexical categories, enabling researchers to look into the deeper meanings of the text. Subsequently, normalization is employed to bring out the various forms of a word to its root level. In this regard, the text is converted into lower case and removed all the punctuation marks. Similarly, all other stop words were also removed from the data. Stop words in natural language processing contribute less to the overall meaning-making process (Munková et al., 2013). The numbers of these words in English are 179. Finally, the text was lemmatized to bring the data back to its root forms. As a result, the text’s variance due to different forms of the word features is avoided.

Latent Dirichlet Allocation

The present study employs the Latent Dirichlet Allocation (LDA) method for data analysis to determine the selected corpus’s hidden meaning. LDA is used in unsupervised machine learning algorithms (Blei et al., 2003; Hoffman et al., 2010). It determines the underlying topics in a more significant chunk of a text which is manually not possible. Besides, LDA encodes the intuition that documents cover only a small set of topics and frequently uses only a small set of keywords (Zhang et al., 2013).

LDA is a robust method to study hidden meaning out of a larger chunk of the textual data under consideration. It deals with unsupervised data. Most qualitative studies such as sociological research, opinion analysis, and media studies benefit from automated topic mining (Nikolenko et al., 2017). It is equally helpful in dealing with large corpora. Brookes and McEnery (2019) refer to the topic modeling approach that is useful to group texts that are truly thematically coherent for discourse analysis. Topic modeling reveals, for example, underlying meaning out of newspaper corpora on a given topic (Uys et al., 2008). Contextual studies from different countries worldwide have been conducted to identify newspapers’ discourses (Åkerlund, 2019; Viola & Verheul, 2020). Besides, topic modeling is a way to understand the framing strategies on the media agencies’ part (Heidenreich et al., 2019; Semetko & Valkenburg, 2000). This study is based on the corpus of Pakistani English newspapers. It uses Latent Dirichlet Allocation (Blei et al., 2003) technique, a robust algorithm to extract topics from a corpus. Hence, the topic modeling can help understand the nature of Pakistan’s economy challenges.

The study extracted 10 groups of topics. Each group consisted of ten keywords. These keywords were analyzed in detail to find out common themes. As a result, these themes helped in getting the overall picture of the text, which summarized the Pakistani economy dynamics. LDA helps to visualize the word clouds of the extracted topics (Bashri & Kusumaningrum, 2017; Ganesan et al., 2015). Furthermore, the data visualization demarcates topics and promotes comparison both within and across latent topics. Likewise, the visualization of topics clarifies the meaning, prevalence, and relation of topics (Sievert & Shirley, 2014).

Theoretical Framework

The theoretical framework of this research is based on grounded theory which proposes that the meaning is grounded in the text. It is a data-driven approach that guides researchers to reach the problem’s fundamental nature (Charmaz, 2014; Morse et al., 2016). It also focuses on both phenomenological and positivist roots. It is a systematic, inductive, and comparative approach for conducting an inquiry (Bryant & Charmaz, 2007). This sort of research does not require the prior assumption of hypotheses or research questions. In the data-driven approach, the text considered for evaluation is the primary source that brings forth the meaning from the text and informs researchers to decide upon it (Birks & Mills, 2015). Thus, using this approach
under the umbrella of grounded theory, the data were retrieved, and then the topic modeling was employed on the data through LDA. Finally, the extracted topics were organized and explained in detail.

**Data Analysis**

The results of the data and the analysis obtained through LDA are summed up in the following table. After analyzing the data, the keywords were extracted. Then, to determine the meanings of keywords, the topics are suggested for more comprehensibility of the analyzed data.

Table 1 explains ten groups of words ranging from 0 to 9 and ten keywords for each, showing various frequencies from higher to lower than the words’ weightage in the written articles. The topics on the above table’s left hand are proposed to reflect a summarized and clear picture of the analyzed data for scanning purposes. Overall, the analyzed data of the table are further explained below:

The first set of keywords in Table 1 suggests a change (0.093) in the existing laws (0.031) for a new economic direction (0.034) in Pakistan. This change in law refers to the policies of the country regarding different sectors. It is a demand (0.027) to study (0.022) the existing structure and put things in order (0.020). The transition from wrong (0.046) policies to new ones may help overcome the debt hike (0.021).

The second set of keywords relates to export (0.073) in textile (0.040). The keywords anti (0.039), corruption (0.042), hold (0.039), and concern (0.030) predict the rise of corruption (0.042) to be the primary concern to improve the textile industry. Anti-corruption reforms and improving business may help to mobilize investment and drive the economy in Pakistan. Exports are a primary source of decreasing the debt of the country. There is a dire need to address corruption in the area of exports and to facilitate the industry. The keyword woman (0.026) reflects women’s participation in the textile industry in Pakistan.

The third set of extracted keywords relates to security preference. The keywords peace (0.050), resources (0.033), curb (0.027), and influence (0.023) refer to measures relating to security. The security preference is linked with the social and political condition of the country to boost the economy.

The fourth set of keywords signals a shift of energy from fuel (0.008) to solar (0.006) mode. The keyword affect (0.057) is the word in the group with the highest weightage. The term effect indicates the transition from fuel to solar can forecast (0.023) a positive change in the country’s economy as the government has to spend a lot of revenue on its energy consumption. The term license (0.006) refers to the facilitating solar energy-based project on the part of the government.

The fifth set of keywords refers to high inflation (0.072) and low GDP (0.016). High inflation is a challenge for the government. There is a need to plan (0.028) and introduce policies (0.041), especially to look after the price hike in food (0.026) commodities. The high inflation results in low GDP, which directly affects the overall economic condition of the country.

The sixth set of keywords relates to cooperation and engagement. The keywords close (0.013), engage (0.000), wayward (0.000) refer to harmony between different country sectors. On the other hand, the terms conflict, willing, lesson, agree, and truce refer to learn a lesson from the past
and resolving conflicts. The need is to decide on a truce to bring the country out of the economic crises.

The seventh set of keywords relates to growth (0.084) and investment (0.051) in the power sector (0.062) to enhance the present state of the economy. The low (0.33) rate (0.028) of energy (0.025) generation is likely to affect growth (0.084). On the other hand, energy investments are likely to raise the power (0.037) capacity, affecting all major growth and investment sectors. The growth and investment sector needs to be boosted as there is a triangle linkage between tax, growth, and investment.

The eighth set of keywords refers to call (0.045) for starting mega (0.030) projects (0.032) to deal with economic (0.040) issues (0.045). Among the big projects, CPEC is the main project. Some political sects protest (0.027) that bilateral issues, including Chinese financing in CPEC, result in a rise (0.025) in debt burden with no transparency.

The ninth set of keywords relates to sustainable democracy. The keywords talk, discuss, youth challenge, and face refer to involve the country’s workforce, especially youth. The solution to this poverty control is a preference to trade in the country and with neighboring states.

The last set of keywords relates to poverty control. To control poverty and economic (0.055) crises, there is a need to boost trade (0.029) in the country. The increase in debt may raise poverty, and people may get poor (0.023). The need of the time is to end (0.020) this crisis. Moreover, poverty is a primary threat to the dwindling economic condition of the country.

In addition to the analyzed data provided in Table 1, the results are further visualized in the following figures for getting a more pretty clear picture in terms of meaning, occurrence, and relation of the topics.

Figure 1 reflects the overall term frequency (blue color) of the extracted topics and the estimated term frequency within the selected topic (red color). The figure visualizes the top 30 most relevant terms of the generated ten groups for comparison both within and across latent topics. The visualization of topics clarifies the meaning, prevalence, and relation of topics.

In addition to topic visualization, the extracted topics through topic modeling are also shown in Figure 2 in form of word clouds of the keywords ranging from 0 to 9.

The keywords extracted through topic modeling using LDA have been shown in Figure 1. The size of keywords in Figure 1 suggests representation of the keywords based on the weightage of words in the specific topic, which helped decide the suggested topic and critical issue.

Furthermore, t-SNE distribution of topics was conducted. A t-SNE visualization method visualizes high-dimensional data by giving each data point a location in a two or three-dimensional map (Van der Maaten & Hinton, 2008).
Figure 3 shows the distribution of topics through t-SNE obtained through LDA. The topics’ distribution shows that the topics are clear and distributed on different continuums predicting clearly the overall size and distribution.

Discussion

The data analysis based on the Pakistani economy shows multiple topics of Pakistani English print media (2015–2020). Generated themes refer to the understanding of the extracted topics regarding the Pakistani economy through words like “change,” “laws,” “peace,” “society,” “textile,” “export,” “solar,” “energy,” “inflation,” “policy,” “GDP,” “investment,” “trade,” “economy,” “challenge,” etc.

The analyzed corpus using LDA presents the statistics of debt dynamics of Pakistan’s economy over the past years (2015–2020) through different generated groups of words. Writers and journalists take on the challenge of highlighting the significant economic issues and deteriorating economy of Pakistan. Their particular focus is on the topics like “laws to change wrong directions,” “security preferences,” “shift from fuel to solar energy,” “low GDP,” etc.

Generated data by print media is modeled by LDA application helps to extract and frame the topics mentioned above in Table 1. Figures and tables clearly show a quantitative description of the data. However, qualitative description is required to understand the changing dynamics of Pakistan’s economy and debt. Improved governance can be observed under the influential role of media. Accordingly, the detailed analysis of the economic news in print media makes economic dynamics clear, and policymakers understand Pakistan’s current challenges and situation. Journalists hit the target of financial debt representation to highlight the current deficits and standing of Pakistan’s economy by critically evaluating the economic system’s loopholes (Adnan et al., 2019).

One of the areas of extracted themes is changing and regulating different economic policies. These policies are uncertain and not durable. One of the reasons for this uncertainty is the lack of consistent policy (Choudhary et al., 2020). As a result, the print media highlights the need to review the existing economic policies and develop a consistent policy to tackle the economic issue. One possible reason may be adopting consistent policies may result in a better economic situation than the situation observed from 2015 to 2020 analyzed in the present study (Mahmood et al., 2008).

The topics generated through different word clusters show that Pakistan’s economy is taking a turn by shifting to solar energy resources instead of sticking to fuel resources, which is an excellent initiative to lower debt to meet world challenges (Khalil & Zaidi, 2014). Generating electricity from renewable resources is an option for Pakistan’s economy to
uplift its status (Amer & Daim, 2011). The severe energy crises in the last decade affected the growth of Pakistan’s economy (Akhtar et al., 2012). It is important to mention that the energy crises is linked to various sectors (i.e., foreign direct investment, textile, and industry) that play role in the economic growth of a country (Latief & Lefen, 2019).

The extracted topics revealed that Pakistan needs to call for starting mega projects in the country to overcome the economic issues. A project like China Pakistan Economic Corridor (CPEC) is mainly referred in the data to be considered a gateway for a flourishing economy in Pakistan. It may be one of the prospects of CPEC in eradicating poverty eradication, growing economy, enhancing power resources, building infrastructure, and giving employment to youth and locals. In this regard, the print media has marketed the potential growth of CPEC projects and their benefits nationally and internationally. Being the fourth pillar of the society, it details the future investment prospects in CPEC and provides stakeholders’ information in the masses. Likewise, Mengal et al. (2018) refer to the importance of CPEC as, “once this project is completed, it will boost Pakistan’s economy, and it will go upwards” (p. 1). Besides, the corpus analysis also shows that Pakistani print media equally raised the voice of some people who have apprehensions about the CPEC project that it would increase Pakistan’s debt due to the shared burden of the economy with China. However, sustainable democracy is an option to eradicate the economic issue of the country (Bibi et al., 2018; Rizvi, 2011). In this regard, Pakistani youth and the rest of the professionals may play a vital role in facing the challenges in the region (Zaidi, 2005b). Peaceful talk and discussion with neighboring countries like India on democratic grounds may also help lessen the country’s economic burden (Wolf, 2016). It may increase trade and employment, which ultimately control poverty. Generally, poverty is a hindrance to the economy’s growth, and free trading is a reduction of debt in a country like Pakistan.

The study underpinned grounded theory which explains that the meaning is grounded in the text. In other words, it adopts a data-driven approach to drive meaning out of a larger volume of text under consideration. The essence of the grounded theory is to come up with an understanding of the matter. By using grounded theory, the present study brings to the fore the challenges and prospects relating to the economy of Pakistan.

Limitations and Future Research Directions

Firstly, the study is limited to one country’s economic conditions, whereas the economic condition involves multiple factors, the country’s internal condition, and external relations with other countries. The present research is limited only to the information available within newspapers articles. Secondly, the data were taken in this study only relies on newspaper articles. The data from other sources relating the situation of the economy may be helpful to triangulate the information based on economic condition of Pakistan.

Conclusion

This study concludes that laws to change economic policies, security preference, textile industry and export, the shift of energy from fuel to solar techniques, inflation, cooperation and engagement, growth, mega projects, sustainable democracy, and poverty control are the main topics extracted from the corpus-based on the economy of Pakistan. This corpus analysis has depicted practical and social issues in declining the debt of Pakistan. With its linguistic tools, topic modeling, and LDA, natural language processing help determine the text’s nature in print media on financial debt and GDP issues. This type of research is not very common in the context of South Asian academic research. Through topic modeling and word clustering, it becomes clearer that representation of facts and figures helps the state revise and align its policies as analyzed and suggested by print media.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Fasih Ahmed https://orcid.org/0000-0002-9357-7073

References

Abadie, A., & Gardeazabal, J. (2008). Terrorism and the world economy. *Global Economic Review, 52*(1), 1–27.

Adnan, M., Ali, A., & Aslam, S. (2019). Economic issues and ethical journalism in Pakistan: Prospects and challenges. *Global Social Sciences Review, 4*(1), 11–22.

Afzal, M. (2009). Population growth and economic development in Pakistan. *The Open Demography Journal, 2*(1), 1–7.

Ahmed, R. N., & Ahmad, K. (2016). Impact of population on economic growth: A case study of Pakistan. *Bulletin of Business Economics, 3*(3), 162–176.

Ahmed, V. (2017). *Pakistan agenda for economic reforms*. Oxford University Press.

Åkerlund, M. (2019). Representations of trans people in Swedish newspapers. *Journalism Studies, 20*(9), 1319–1338.

Akhtar, S., Javed, B., Maryam, A., & Sadia, H. (2012). Relationship between financial leverage and financial performance: Evidence from fuel & energy sector of Pakistan. *European Journal of Business and Management, 4*(11), 7–17.
Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences, 12*(16), 4102–4107.

Ali, S., Ali, A., & Amin, A. (2013). The impact of population growth on economic development in Pakistan. *Middle-East Journal of Scientific Research, 18*(4), 483–491.

Amer, M., & Daim, T. U. (2011). Selection of renewable energy technologies for a developing county: A case of Pakistan. *Energy for Sustainable Development, 15*(4), 420–435.

Amjad, R., & Burki, S. J. (2015). Pakistan: Moving the economy forward. Cambridge University Press.

Bashri, M. F., & Kusumaningrum, R. (2017). Challenges for democracy in Pakistan. *Global Political Review, 3*(1), 66–75.

Bibri, F., Jameel, S., & Jalal, S. U. (2018). What is democracy? Challenges for democracy in Pakistan. *Global Political Review, 3*(1), 66–75.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(2003), 993–1022.

Brookes, G., & McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies, 21*(1), 3–21.

Bryant, A., & Charmaz, K. (2007). *The Sage handbook of grounded theory*. Sage.

Butt, K. M., & Butt, A. A. (2015). Impact of CPEC on regional and extra-regional actors. *The Journal of Political Science, 33*, 23.

Chandukala, S. R., Kim, J., Otter, T., & Allenby, G. M. (2008). *Choice models in marketing: Economic assumptions, challenges and trends*. Now Publishers Inc.

Charmaz, K. (2014). *Constructing grounded theory*. sage.

Choudhary, M. A., Pasha, F., & Waheed, M. (2020). *Constructing grounded theory*. sage.

Choudhary, M. A., Pasha, F., & Waheed, M. (2020). Measuring economic policy uncertainty in Pakistan. *Munich Personal RePEc Archive.*

Dao, M. Q. (2012). Population and economic growth in developing countries. *International Journal of Academic Research in Business, 2*(1), 6.

Denny, M., & Spirling, A. (2017, September 27). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *When It Misleads, and What to Do about It, Political Analysis, 26*(2), 168–189.

Easterly, W. (2001). The political economy of growth without development: A case study of Pakistan. In: Dani Rodrik (Ed.), *Search of prosperity: Analytic narratives on economic growth* (pp. 439–472). Princeton University Press, 2012.

Fatima, M., Latif, M., Chughtai, S. F., Hussain, N., & Aslam, S. (2014). Terrorism and its impact on economic growth: Evidence from Pakistan and India. *Middle-East Journal of Scientific Research, 22*(7), 1033–1043.

Feehey, G., & Alam, I. (2003). New estimates and projections of population growth in Pakistan. *Population Development Review, 29*(3), 483–492.

Ganesan, A., Brantley, K., Pan, S., & Chen, J. (2015). Ldaexplore: Visualizing topic models generated using latent dirichlet allocation. *arXiv preprint arXiv:1507.06593.*

Gul, S., Irshad, F., & Zaman, K. (2011). Factors affecting bank profitability in Pakistan. *The Romanian Economic Journal, 14*(39), 61–87.

Hamid, N., Nabi, I., Nasim, A., & Mundial, B. (1990). *Trade, exchange rate, and agricultural pricing policies in Pakistan* (Vol. 67). Citeseer.

Heidenreich, T., Lind, F., Eberl, J.-M., & Boomgaarden, H. G. (2019). Media framing dynamics of the ‘European refugee crisis’: A comparative topic modelling approach. *Journal of Refugee Studies, 32*(Special Issue 1), I172–I182.

Hoffman, M., Bach, F. R., & Blei, D. M. (2010). Online learning for latent dirichlet allocation [Paper presented]. The advances in neural information processing systems, 23.

Hongming, X., Ahmed, B., Hussain, A., Rehman, A., Ullah, I., & Khan, F. U. (2020). Sustainability reporting and firm performance: The demonstration of Pakistani firms. *SAGE Open, 10*(3), 2158244020953180.

Hussain, E., & Rao, M. F. (2020). China–Pakistan economic cooperation: The case of special economic zones (SEZs). *Fudan Journal of the Humanities and Social Sciences, 13*(4), 453–472.

Hussain, I. (2000). *Pakistan: The economy of an elitist state*. OUP Catalogue.

Hussain, I., Reddy, Y., & Kamil, M. S. (2018). Governing the ungovernable. Oxford University Press.

Hyder, S., Akram, N., & Patta, I. U. H. J. P. B. R. (2015). Impact of terrorism on economic development in Pakistan. *Pakistan Business Review, 839*(1), 704–722.

Jiang, T., Guo, Q., Chen, S., & Yang, J. (2019). What prompts users to click on news headlines? Evidence from unobtrusive data analysis. *Aslib Journal of Information Management, 72*(1), 49–66.

Khalil, H. B., & Zaidi, S. J. H. (2014). Energy crisis and potential of solar energy in Pakistan. *Renewable and Sustainable Energy Reviews, 31*, 194–201.

Khan, S., & Liu, G. (2019). The China–Pakistan economic corridor (CPEC): Challenges and prospects. *Area Development and Policy, 4*(4), 466–473. https://doi.org/10.1080/23792949.2018.1534549

Koh, W. T. H. (2007). Terrorism and its impact on economic growth and technological innovation. *Technological Forecasting, 74*(2), 129–138.

Latief, R., & Lefen, L. (2019). Foreign direct investment in the power and energy sector, energy consumption, and economic growth: Empirical evidence from pakistan. *Sustainability, 11*(1), 192.

Lewis, M. (2003). *LexisNexis™ Academic*. New Republic.

Mahmood, T., Rehman, H.-u., & Rauf, S.-A (2008). Evaluation of macro economic policies of Pakistan (1950–2008). *Journal of Political Studies, 17*, 57–75.

Masakure, O., Cranfield, J., & Henson, S. (2011). Factors affecting the incidence and intensity of standards certification evidence from exporting firms in Pakistan. *Applied Economics, 43*(8), 901–915.

Mengal, T. M., Hussain, S., Taj, M. K., Liang, Z., & Taj, I. (2018). Role of print media in China-Pakistan economic corridor as ungovernable. *Oxford University Press.*

Mirjat, N. H., Uqaili, M. A., Harijan, K., Valasai, G. D., Shaikh, F., & Waris, M. (2017). A review of energy and power planning and development: A case study of Pakistan. In: Dani Rodrik (Ed.), *Macro economic policies of Pakistan (1950–2008).* Journal of Development Policy, 57–75.
policies of Pakistan. Renewable Sustainable Energy Reviews, 79, 110–127.
Morse, J. M., Stern, P. N., Corbin, J., Bowers, B., Charmaz, K., & Clarke, A. E. (2016). Developing grounded theory: The second generation (Vol. 3). Routledge.
Munková, D., Munk, M., & Vozár, M. (2013). Influence of stopwords removal on sequence patterns identification within comparable corpora [Paper presentation]. In International Conference on ICT Innovations (pp. 67–76). Springer, Heidelberg.
Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modelling for qualitative studies. Journal of Information Science, 43(1), 88–102.
Peterson, E. W. F. (2017). The role of population in economic growth. SAGE Open, 7(4), 2158244017736094. https://doi.org/10.1177/2158244017736094
Rehman, A., Jingdong, L., & Hussain, I. (2015). The province-wise literacy rate in Pakistan and its impact on the economy. Social Sciences Pacific Science Review B: Humanities, 1(3), 140–144.
Rizvi, H. A. (2011). Democracy in Pakistan [Paper presentation]. In Panorama: Insights into Asian and European Affairs is a series of occasional papers published by the Konrad-Adenauer-Stiftung’s Regional Programme Political Dialogue Asia/Singapore (p. 117).
Saboor, A., Khan, A. U., Hussain, A., Ali, I., & Mahmood, K. (2015). Multidimensional deprivations in Pakistan: Regional variations and temporal shifts. The Quarterly Review of Economics, 56, 57–67.
Semetko, H., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. Journal of Communication, 50(2), 93–109.
Shahbaz, M., Shabbir, M. S., Malik, M. N., & Wolters, M. E. (2013). An analysis of a causal relationship between economic growth and terrorism in Pakistan. Economic Modelling, 35, 21–29.
Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics [Paper presentation]. In Proceedings of the workshop on interactive language learning, visualization, and interfaces, Baltimore, MD, USA, pp. 63–70.
Tognini-Bonelli, E. (2001). Corpus linguistics at work (Vol. 6). John Benjamins Publishing.