BIG DATA APPLICATIONS IN MONETARY POLICY AND FINANCIAL STABILITY FOR SEACEN MEMBER ECONOMIES: THE CASE OF BANK INDONESIA

Sahminan Sahminan, Tri Yanuarti, Ayi Supriyadi,
Pahmi Utamaraja Ginting, Jati Waluyo, Dwi Oktaviyanti,
Rieska Indah Astuti, Risa Latul Ayuni, Muhammad Zikri,
Irfan Sampe, Mohammad Abdul Jabbar, Mohammad Khoyrul Hidayat,
Amin Endah Sulistiawati, Aryo Sasongko

The SEACEN Centre

The South East Asian Central Banks (SEACEN)
Research and Training Centre
198201000672 (80416-M)

BIG DATA APPLICATIONS IN MONETARY POLICY AND FINANCIAL STABILITY FOR SEACEN MEMBER ECONOMIES: THE CASE OF BANK INDONESIA

The SEACEN Centre

The South East Asian Central Banks (SEACEN) Research and Training Centre
198201000672 (80416-M)

FOREWORD

In 2006, British mathematician Clive Humby asserted that 'data is the new oil', and *Forbes* magazine proclaimed the 'Age of Big Data' in 2012. The relentless rise in digitalisation has been driven by developments on the supply as well as the demand side. On the supply side, the mainly big technology (BigTech) companies doing the digital data collection realised that data, like oil, was not useful in its raw state, but needed to be refined, processed and turned into something useful to be monetised. On the demand side, the Great Financial Crisis of 2008-2009 underscored the need for central banks and regulatory authorities to have access to more disaggregated and timely data about the health of the macro-financial system. The COVID-19 pandemic in 2020 only served to further underline the benefits of having access to non-traditional, high-frequency data for a real-time assessment of the economy.

By Big Data we generally understand large datasets that in many cases are non-traditional and unstructured, although large, structured datasets, such as those arising from administrative sources or those based on regulatory reporting requirements, also qualify as Big Data. One identifying aspect is that the data sets are much larger than the traditional ones collected by national statistical offices or the central bank itself. For most central banks, though, Big Data tends to complement traditional data sources with non-traditional ones to inform policy decisions. The latter substitute for a lack of reliable official statistics or help in cases where these statistics are published with a sizeable lag. As the types of data gathered and processed by central banks have continuously changed and expanded over the past two decades, so have the tools to analyse this information. In fact, new technological capabilities have led to major changes in the type of data that central banks collect and subsequently the way these data are used for policymaking.

The use of Big Data, and the associated data analytical techniques, is inexorably moving forward in central banking, including in many SEACEN member central banks and monetary authorities. In 2020, over 80 per cent of central banks surveyed by the Bank for International Settlement's Irving Fisher Committee on Central Bank Statistics (BIS-IFC) reported that they were using Big Data, an increase from some 30 per cent in 2015. More specifically, Asian central banks were some of the most enthusiastic adopters of Big Data and the associated analytical toolkit, such as natural language processing, nowcasting/monitoring exercises, and applications to extract economy insights as well as SupTech/RegTech solutions. In general, central banks and supervisory authorities in Asia already use Big Data sources such as machine learning (ML) extensively for research purposes, namely, to inform and assist monetary policy decisions, facilitate their statistical compilation work, and support their regulatory supervisory tasks.

One of the leading proponents of Big Data and the associated data analytics toolkit amongst SEACEN members is Bank Indonesia (BI) and we are grateful to BI for having contributed this survey on 'Big Data Applications in Monetary Policy and Financial Stability for SEACEN Member Economies: The Case of Bank Indonesia' to our research project. Just like many other central banks around the world, BI has been increasingly using granular Big Data for research, policy formulation and decision-making. Since 2015, BI has developed and utilised Big Data analytics in 40 pilot projects to strengthen the process of formulating monetary, macroprudential and payment system policies.

The research paper highlights several important aspects of Big Data and the data analytical toolkit, starting with the many possible uses cases Big Data can be used for and the ever-increasing capability of data analytical techniques. Big data at BI, just as in many other central banks, is used in a variety of ways, including research, monetary policy, financial stability and regulation and supervision.

But the undoubted analytical opportunities must be seen in light of challenges, such as technical or operational, human resource, product quality (and assurance) and legal aspects. More specifically, for central banks starting off in this area, issues to consider are data availability, collection, quality, sampling, representativeness, security, processing and storage, privacy concerns and data governance. In addition, central banks face high up-front costs that go hand-in-hand with data availability, legal uncertainty around data privacy and confidentiality. Finally, there are constraints in setting up an adequate IT infrastructure and in developing the necessary human capital. Frequently, this involves tough questions about prioritising projects.

When it comes to Big Data, a lot of data creation has migrated from the public to the private sector, which raises the issue of data sharing and whether central banks should have to pay BigTech firms for the data, or whether the latter should be compelled to provide the data to public authorities for free. Analysis of such data raises communication challenges, not only within the central bank but also in terms of the external communication of the monetary and financial stability authority. Interestingly, economic agents adjust to new technologies as well, in the sense that public statements and filings now frequently exclude words that trigger negative sentiment or measure uncertainty in natural language/textual processing.

Evidence so far also underscores the important role of co-operation among public authorities to improve central banks' ability to collect, store and analyse Big Data. On the topic of co-operation, the BIS-IFC survey also highlights the important role of central bank co-operation, which is expected to foster central banks' use of Big Data. More specifically, sharing knowledge among those institutions that have developed specific expertise that can be re-used in other jurisdictions was identified as important. The SEACEN's research paper that collects and showcases successful Big Data projects should be seen in the spirit of facilitating the sharing of experience. It is our hope that SEACEN member central banks and monetary authorities will use these innovative news forms of data and data techniques to inform and assist monetary policy and financial stability going forward.

Mangal Goswami
Executive Director
The SEACEN Centre

ABSTRACT

BIG DATA APPLICATIONS IN MONETARY POLICY AND FINANCIAL STABILITY FOR SEACEN MEMBER ECONOMIES: THE CASE OF BANK INDONESIA

Sahminan Sahminan[1], Tri Yanuarti[1], Ayi Supriyadi[1], Pahmi Utamaraja Ginting[1], Jati Waluyo[2], Dwi Oktaviyanti[2], Rieska Indah Astuti[2], Risa Latul Ayuni[2], Muhammad Zikri[3], Irfan Sampe[3], Mohammad Abdul Jabbar[3], Mohammad Khoyrul Hidayat[3], Amin Endah Sulistiawati[3], Aryo Sasongko[4]

Big Data analytics play an important role in today's digital era, including the support of the formulation of both monetary and macroprudential central bank policies. In addition, the central bank also utilises Big Data for the supervision and regulation of the financial system, such as the application of supervisory technology (SupTech) and regulatory technology (RegTech). The COVID-19 pandemic, which rapidly changed the behaviour of economic agents and companies, has accelerated the important role of Big Data analytics. In the Asia-Pacific region, some central banks have taken advantage of Big Data analytics, particularly for macroeconomic projections and financial supervision. Bank Indonesia, as a member of SEACEN, has developed and utilised Big Data analytics through 40 pilot projects since 2015 to strengthen the process of formulating monetary, macroprudential and payment system policies. But the development of Big Data analytics is still faced with a number of challenges, such as the availability and quality of the infrastructure, a lack of experts and unclear legal requirements, as well as the security of its use. Going forward, the development of Big Data analytics needs to be directed towards sharpening granular data use cases, improving data quality and coverage, increasing human resource capabilities, expanding data acquisition collaboration with various parties, more forward-looking analysis coverage and developing new indicators.

1. Economic and Monetary Policy Department, Bank Indonesia.
2. Macroprudential Policy Department, Bank Indonesia.
3. Statistics Department, Bank Indonesia.
4. Bank Indonesia Institute.

TABLE OF CONTENTS

BIG DATA APPLICATIONS IN MONETARY POLICY AND FINANCIAL STABILITY FOR SEACEN MEMBER ECONOMIES: THE CASE OF BANK INDONESIA

List of Tables

List of Figures

BIG DATA APPLICATIONS IN MONETARY POLICY AND FINANCIAL STABILITY FOR SEACEN MEMBER ECONOMIES: THE CASE OF BANK INDONESIA

Sahminan Sahminan, Tri Yanuarti, Ayi Supriyadi, Pahmi Utamaraja Ginting, Jati Waluyo, Dwi Oktaviyanti, Rieska Indah Astuti, Risa Latul Ayuni, Muhammad Zikri, Irfan Sampe, Mohammad Abdul Jabbar, Mohammad Khoyrul Hidayat, Amin Endah Sulistiawati, Aryo Sasongko

1. Introduction

1.1 The Important Role of Big Data in Supporting Central Bank Policy Formulation

Big Data analytics play an essential role in today's digital era, including in support of a central bank's formulation of its monetary and macroprudential policies. Empirical findings by the European Commission (2016) in the context of public policy formulation show that the use of Big Data analytics does not only improve the quality, quantity and type of data, but also has great potential to influence the entire cycle of public policy formulation to make it more efficient and effective. A study by the Bank for International Settlements (BIS) (Doerr et al., 2021) indicates that most central banks have utilised Big Data, including machine learning (ML), in various areas, such as research and formulation of monetary policy and financial system stability. In addition, the central bank also utilises Big Data for supervision and regulation of the financial system, such as applying supervisory technology (SupTech) and regulatory technology (RegTech). But several problems arise in utilising Big Data analytics, such as the quality of data and sampling appropriation, as well as legal questions related to the privacy and confidentiality of data.

Big Data analytics became increasingly important with the onset of the COVID-19 pandemic in early 2020. The COVID-19 pandemic has rapidly changed the behaviour of economic agents and companies, both in the real and the financial sector, which presents an obstacle for economic policymakers in trying to understand the factors that underlie economic development. Traditional statistical data, as well as existing models, are often inadequate to support the decision-making process because they are based on low-frequency data that are published with a (sizeable) lag. Meanwhile, the impact of the COVID-19 episode underscores the benefits for carrying out analysis using higher-frequency, real-time and granular data. This has served to further encourage efforts amongst policymakers to develop and utilise Big Data analytics.

1.2 The Use of Big Data by Central Banks in the Asia-Pacific Region

Some central banks have taken advantage of Big Data analytics already. Results from the Executives' Meeting of East Asia-Pacific Central Banks (EMEAP) survey in 2020, show that some central banks in the region have a great desire to continue to develop alternative data using Big Data analytics. This is especially so in the Asia-Pacific region, based on the last survey from IFC-BIS in 2020 as interest in using Big Data analytics tends to be higher for central banks in the Asia-Pacific region compared to central banks in other regions, with a particular focus on the development of natural language processing, now-casting, more granular use of financial data and SupTech/RegTech (Cornelli et al., 2022).

1.3 Big Data Development Challenges and the Importance of Identifying Big Data Development Levels in SEACEN Member Central Banks

Several challenges are still being faced in the development of Big Data analytics, including the availability of infrastructure, both on the hardware and software side, that can be relied upon to process Big Data; the limited number of experts (data scientists) who have the required skills to be able to perform Big Data analytics; the process of preparing the data and its reconciliation with data published by other institutions; data governance; legal aspects; and, the ethics involved in using private data and issues relevant to cyber security. In line with such developments, The South East Asian Central Banks (SEACEN) Research and Training Centre is also interested in developing Big Data analytics in the central banks of its member economies. The level of progress and different information technology infrastructure between central banks impacts the pattern and level of adoption and development of Big Data analytics. Therefore, as an initial step, it is necessary to conduct an in-depth identification of the level of development of Big Data analytics in various central banks of SEACEN member economies, especially for their use in supporting monetary and macroprudential policy formulation. From the identification results, lessons learnt and prerequisites that are important for developing Big Data analytics in the region can be identified. In addition, the results of such an analysis can also be used as a basis for developing collaborative initiatives between SEACEN member central banks to use Big Data analytics in the future.

1.4 The Role of Bank Indonesia in Encouraging the Development of Big Data

As a SEACEN member, Bank Indonesia has been encouraging the development and utilisation of Big Data analytics since 2015. This has been done through various pilot projects to strengthen the process of formulating monetary, macroprudential and payment system policies. Such efforts are in line with the achievement of Bank Indonesia's 2025 vision "To become the foremost digital central bank that creates a tangible contribution to the national economy and the best amongst emerging market countries towards an Advanced Indonesia". In general, the use of Big Data Analytics to strengthen the policy-making process at Bank Indonesia is directed at four areas, which are: (i) overcoming data issues by providing new data/indicators more quickly; (ii) supporting the analysis of macroeconomic developments and the behaviour of market and financial system actors; (iii) the establishment of network analysis, especially the analysis of linkages between economic actors, for the monitoring of, *inter alia*, systemic risk; and (iv) the assessment of public perceptions and expectations of Bank Indonesia policies. So far, the data and analysis produced by Big Data analytics have been used by various departments at Bank Indonesia to support the formulation of monetary policy, macroprudential, payment systems and supervision. In the future, several challenges in the development of Big Data analytics will still need to be tackled, such that the considerable opportunities for strengthening the supporting role in the formulation of Bank Indonesia policies can be attained.

1.5 Contribution of Mapping the Utilisation of Big Data at Bank Indonesia

Based on this background, this study seeks to map the use of Big Data analytics at Bank Indonesia. This includes various Big Data analytics use cases developed by Bank Indonesia in various areas, both monetary and macroprudential, and their usage in supporting policy formulation at Bank Indonesia.

This paper consists of seven parts. The second part presents a literature study on the concepts and benefits of Big Data, particularly in central banks. The development of Big Data at Bank Indonesia is outlined in Part 3. Part 4 describes the use of Big Data in supporting the formulation of monetary policy at Bank Indonesia. Part 5 describes the use of Big Data in supporting macroprudential policy formulation. Part 6 describes the funding challenges for the future development of Big Data. Finally, in Chapter 7, several conclusions and policy implications are discussed.

2 Review of the Literature

2.2 Big Data Concepts and Benefits

Big Data cannot be separated from data mining, ML and text mining. Data mining is a tool that analyses data in Big Data sets to obtain hidden information contained therein (Hassani et al., 2018). Bholat (2015) outlines that Big Data has three characteristics, which are high volume (detailed information), high-speed velocity (frequently updated) and wide variety (structured and unstructured data, such as non-numeric numeric data, e.g., text and social media videos). Asia-Pacific central banks define Big Data as a data repository that includes large volumes of non-traditional, traditional and unstructured as well as structured data (Cornelli et al., 2022).

Machine learning is part of an artificial intelligence (AI) computer programme that performs several tasks and can learn from its experiences (Ray, 2019). Machine design allows pre-programmed statistical models and algorithms to make specific decisions and predictions/estimates based on data without an explicit programming code (Mahesh, 2020). The more sophisticated ML is, the more complex the machine's ability to perform such tasks. Machine learning applications can help get work done in a variety of fields, with applications that include classification, regression and clustering.

Machine learning uses algorithms for classification and grouping. Programmers design ML applications that use algorithms that will then recognise the structure and features of the findings to identify new patterns or data classes. There are two classes of ML algorithms that are often used, i.e., supervised learning (for example, support vector machines (SVM) and naive Bayes) and unsupervised learning (for example, principal component analysis (PCA) and *k*-means) (Mahesh, 2020).

Supervised learning algorithm is a data mining procedure that maps input to output based on examples of input-output pairs with external assistance or human intervention. The input data group consists of training data and test data. Training data helps the algorithm to conclude a method. The algorithm uses several methods to study different patterns of training data, find and assign outcome variables, and then receive clarification from supervisors/external parties. Next, the algorithm applies the result method to the test data set for prediction or classification.

More specifically, in terms of the supervised learning algorithms, SVM uses a kernel trick, which is to map the input data to a higher dimension. The addition of dimensions could use linear or non-linear equations. The algorithm will determine the appropriate line or field border of the data class/group in the high-dimensional space. Naive Bayes is an algorithm based on Bayes' Theorem with the assumption of independence between predictors of input data. Algorithms classify data based on proximity normal probability density function/conditional probability density function.

Unsupervised learning algorithms, on the other hand, have no external intervention, and unlike supervised learning above, there are no right or wrong answers and no external supervisor. Some examples of unsupervised learning algorithms are PCA and *k*-means. In contrast to other ML algorithms, the PCA algorithm serves to reduce complex dimensions and data visualisation to a simple one. This algorithm only works with data that has been grouped according to dimensions. *K*-means is an algorithm that groups data into as many as *k* clusters according to the proximity of the mean value. Cluster division could use the following four methods: Euclidean distance, Hartigan-Wong method, global optimisation and metaheuristics.

Text mining is an input process that comes from a large collection of online texts to find new facts and trends about something discussed (Bruno, 2016). This system feature can calculate the word frequency distribution of documents, explore changes in sentiment and the direction of sentiment, evaluate changes in the level of readability and formality of the text, and try to measure the level of popularity of documents on web pages. Some processes of text mining are quantitative and qualitative text analyses which aim to interpret and analyse all possible meanings embedded in the text (Benchimol et al., 2022). Quantitative text analysis uses simple computational methods, i.e., counting keywords or text that contains certain concepts. Qualitative analysis or semantic network analysis is about finding relationships between words in a text, giving an idea of the structure of sentences and their meanings. The weakness of this method is that it cannot take into account the underlying meaning of the text.

Some analytical text methods are word counting algorithms, relative frequency, latent semantic analysis (LSA) and topic models. The word counting algorithm is popular because of its simplicity, which is to find the similarity of words in a sentence with a standard dictionary. This algorithm mostly uses the Harvard Psycho-Sociological Dictionary word classification. Meanwhile, the relative frequency algorithm is the division of the number of certain important words against the number of other important words in a document to reveal information about the text. This method, classified as a semi-supervised algorithm, can make an inference of the overall sentiment of the document by finding the relative words of the text group and calculating the relative frequency of the underlying words. LSA is the evaluation of a document to find the meaning or concept underlying the document and can be used in many contexts and for many things. LSA can create a table of terms from a document, and the table becomes a reduced vector space that reflects the semantic structure. The topic model is a method for classifying and mapping text into a certain number of topics. For the most part, topic modelling is a set of algorithms that automatically find and classify word patterns that repeat themselves across a collection of documents. One of them is the Latent Dirichlet Allocation (LDA) algorithm, which classifies words into several topics. Each type of document can have several topics. Topic classification uses unsupervised learning which does not have priority in choosing topics.

2.2 Big Data Analysis Methods in Various Central Banks/International Institutions

According to Doerr et al. (2021), about 85 per cent of central banks regularly use Big Data and similar approaches to assist their increasing policy activities. Big Data is significantly used by the central banks of developed countries. This shows the rapid progress of ML and Big Data capabilities in central bank analytics. Central banks have implemented some tools of Big Data, such as text mining and policy transmission observation.

Studies in economics and finance have widely used text mining. Macroeconomic indices such as interest rates, the Consumer Price Index (CPI) and consumer confidence are influenced by monetary policy communication using the narrative method in the news media. The wider use of Big Data and ML techniques has encouraged the development of text mining algorithms. Peoples' names, topics and sentiments can all be automatically "read" and "extracted" from text using this algorithm. With text mining, data sources from text can be used more widely and much faster than manual work (Wibisono, 2021).

Several central banks monitor their monetary policy transmission using Big Data to support their economic analysis needs. Lucca and Trebbi (2009) initiated the application of a web search index to measure the outcome of changes in monetary policy. They predicted the Federal Reserve's future interest rate decisions based on word searches on Google using statements published by the Federal Open Market Committee (FOMC). In addition, they found that communication made the monetary transmission effect on bond's long-term yields quicker than short-term rates. Wohlfarth (2018) finds that the transmission of monetary policy through global channels changes the value of money and capital markets. Therefore, Wohlfarth (2018) suggests a novel index for policy attention based on daily Google Trends data to assess the spillover of volatility in European and US financial markets.

The ECB presents several good examples of Big Data development by leveraging ML techniques to help formulate policies. Machine learning at the Directorate General-Statistics (DG-S) is used for collecting micro data, especially interbank transactions, to find alternative policy responses (ECB, 2019). The ML system consists of the following modules:

1. **MMSR**. The *Money Markets Statistical Reporting* (MMSR) dataset illustrates the application of Information Technology for the collection, storage, processing, compilation and dissemination of money market data originating from euro-area credit institutions. Based on European Union (EU) Regulation No 1333/2014 ECB of 26 November 2014 concerning statistics on money markets (ECB/2014/48), the main objective of the MMSR is to provide comprehensive, detailed and harmonised information on money market statistics in the Euro area. In 2019, the ECB (2019) collected daily information on money market information from 52 major euro-area banks consisting of 45,000 transactions per day. The ECB and national central bank (NCB) daily results are available in Extensible Markup Language (XML) format (in accordance with ISO 20022) and *via* secure transmission channels (ECB, 2017).

2. **EMIR Data**. EMIR Data is a database containing all derivative trading transactions in the EU, with two-sided reporting for buyers and sellers. Between 20 and 100 million transactions per day have been collected since 2014, with 80 to 120 data attributes per transaction (Boumghar, 2019). The term is derived from the European Market Infrastructure Regulation (EMIR). The European Union Regulation No 648/2012 on over-the-counter (OTC) derivatives, central counterparties and trade repositories (EMIR) was officially issued by the EU on 27 July 2012, and came into force on 16 August 2012 (EC, 2016). The regulation establishes rules regarding OTC derivatives, central counterparties, and trading repositories. EMIR strives to promote transparency and standardisation in the derivatives market as well as reduce systemic risk.

3. **AnaCredit**. The *Analytical Credit Dataset (*AnaCredit) is a multipurpose database containing information on credit loans distributed by credit institutions to companies and other legal entities. On 18 May 2016, the ECB Regulatory Board adopted Regulation ECB/2016/13 on the collection of granular credit and credit risk data (AnaCredit) from the European System of Central Bank (ESCB) database (Israël et al., 2017). In 2020, AnaCredit contained detailed information on 60 million individual euro-area bank loans to legal entities (ECB, 2020) and nearly 100 different information attributes covering various aspects of credit exposure, such as amount owed, maturity, interest rate, collateral or guarantee, information about the counterparty, and so on. This information supports the ECB in risk management, banking and financial system stability and collateral management from the Eurosystem, for example asset pricing, credit risk assessment and haircuts, macroprudential analysis and quantitative risk assessment, especially in the context of macro stress testing.

4. **SUP Data**. The *Supervisory Banking Statistic* (SUP) data is a system that produces credible quality data on banking financial positions, risk, liquidity and bank leverage from two sources, which are the Implementing Technical Standards (ITS) and Short-Term Exercises (STE). The ITS and STE provide detailed reports on the financial positions, risk, liquidity and leverage of banks for an estimated total of 5,660 banks (Trezecio and Sánchez, 2019).

5. **RIAD.** The *Register of Institutions and Affiliates Database* (RIAD) is a database containing a list of organisational (financial) units managed by the ECB's DG-Statistics and accessible by all members of the European System of Central Bank (ESCB). RIAD managers also try to ensure a complete and homogeneous data population so that they can provide complete, timely and accurate statistical reports (Israël, 2017). Some examples of ML use cases in DG-Statistics include:

    a. The Anomaly Detector/Outlier Detection where the AnaCredit module is a tool to validate non-conforming credit data which cannot be explained by ordinary statistical means.

    b. The classification of transaction data by checking and/or matching/pairing money market data using the MMSR module and derivative data from the EMIR Data module.

c. The evaluation of data credibility and quality using a special expert system for data validation.

d. The recovery of missing data using the ML algorithm during forecasting, backcasting, and interpolation.

e. The relationship between multiple records of the same entity in different databases and calibration of missing data/data integrity by matching data between MMSR and RIAD.

In addition to the ECB, the Central Bank of the Republic of North Macedonia also analyses credit risk by developing data science and ML techniques. This system processes existing data, i.e., credit data from the central bank. The results of the risk analysis are open to domestic commercial banks to ascertain the credit history of each borrower in the country. Doko et al. (2021) compared five ML models to classify credit risk data, i.e., logistic regression, decision tree, random forest, SVM and neural network. They evaluated five models using different machine learning metrics and found that the decision tree model yielded the best estimation accuracy. This model can calculate risk based on unbalanced data with and without scale, followed by random forest and linear regression.

Several central banks use text mining to analyse perceptions of monetary policy, for example the Central Bank of Nigeria (CBN) and the Central Bank of the Republic of Turkey (CBRT). While the CBN announces the results of its interest rate decision after a meeting of the monetary policy committee (MPC) just like almost all other central banks, it substantially improved its communications year over year, implying increased transparency of monetary policy. Towards that end, CBN used text mining to check the readability, sentiment and topic of policy documents. It found that the word and sentence structures of policy news became more complex, thus reducing their readability. In terms of monetary policy sentiment, it was not uncommon to find values of -10.5 per cent, which reflected the level of policy uncertainty faced by the MPC during the sample period (Tumala and Omotosho, 2019).

The CBRT applies text mining and analysis of variance in the processing of monthly price developments reports (MPDR). The publication of the MPDR occurs a day after the announcement of CBRT statistics in Turkish and English. Research by Eskici and Koçak (2018) found two things: one, noun groups and the most widely used nouns between the 2006 and 2018 reports, and second, proof of the MPDR's consistency with the core inflation rate. The first finding is a list of words that can indicate a measure of core inflation and some sectors, such as services or hotels and accommodation, as well as types of goods such as durable or non-durable goods. The word "increase" was one of the most frequently circulated words, an expected finding due to consistent inflation over ten years. In the second part, MPDR uses a three-way/three-cluster variance model for the annual CPI inflation rate, i.e., trend, seasonal and cluster. In short, statistics on the assumption of annual CPI inflation figures are only explained by two clusters: trend and seasonality. The MPDR yields more significant inflation figures than the conventional model, implying the MPDR is consistent with the annual CPI figures.

2.3 Constraints/Challenges Faced by Central Banks/International Institutions in Big Data Development

Participants in the 2014 conference on “Big Data and Central Banks”, organised by the Bank of England’s Centre for Central Banking Studies (CCBS) raised three obstacles to implementing Big Data, which include granular data, legal considerations and new methods of data processing and analysis. The first constraint relates to the type of granular data handled by the central bank, not aggregated data. Rather paradoxically, this extensive use of micro data can make it easier for the central bank to find large/macro patterns. But analysing granular data can be risky when the rights of using and processing granular data are uneven or managed by each end user without mutual access, thus potentially causing inconsistencies and inefficiencies in the use of Big Data. One way to prevent this is to harmonise and implement a standard definition of granular data rights.

The second obstacle is that detailed and granular data has strict legal requirements and conditions. For example, data on supplier and buyer partners, and prices listed on contracts, could be sensitive. Web pages may contain potentially sensitive data, and their use may create ethical and privacy concerns. Minor legal provisions may have significant and systematic consequences. Such legal barriers could limit access to data collected by the central bank, for example, data collected for surveillance. But data owners (e.g., central banks) can overcome such regulatory limits for a secondary purpose, i.e., that there are access rights for monetary economists at the central bank. In this regard, the central bank must relax all legal restrictions on the use of Big Data to support policy formulation. In addition, the central bank should also have more multilateral agreements that allow cross-border data sharing between regulators.

The last obstacle pertinent to new approaches/methods of data processing and analysis is the change in the habits of central bank processing and analysis from deduction (i.e., discussion based on concepts or theories) to induction (i.e., discussion based on data) because of the implementation of Big Data. Another change is the transformation of statistical data, i.e., the representation of groups of data from the calculation of the mean and standard deviation into more details for specific individuals.

The BIS’s survey of 52 Irving Fischer Committee member central banks in 2020 regarding the implementation of Big Data and ML operations (Doerr et al., 2021) highlighted four issues relevant to Big Data and ML applications at the central bank. These issues concern the definition of Big Data by the central bank, the level of implementation of Big Data, the area of application of Big Data at the central bank, and the challenges of using Big Data at the central bank. On the first issue, the central bank defines Big Data in general, especially the character of “variety”. Many characteristics of central bank data are unstructured, originating from administrative activities and regulatory reporting compliance.

Second, the implementation of Big Data applications in central banks has increased in recent years. Big Data discussions focus on various topics, especially on the availability of Big Data and the tools to process, store and analyse it. The results of the 2015 central bank survey on the topic of Big Data showed 30 per cent of respondents actively discussing the topic at their respective central banks. In the 2020 survey by the BIS, this percentage had risen to 80 per cent, with 60 per cent of respondents reporting a strong interest in Big Data topics, including senior policy makers (Doerr et al., 2021; Cornelli et al., 2022). Third, the areas where Big Data techniques were implemented within the central bank included economic research (with some 70 per cent of respondents) and supporting policy decisions (some 40 per cent of respondents). Several institutions use Big Data as support for financial and monetary system stability, supervisory technology and regulatory technology applications (Doerr et al., 2021; Cornelli et al., 2022).

Finally, the new challenges of implementing Big Data in central banks include data cleansing, data representation and data matching. Data cleansing is cleaning raw data such as text data obtained from newspapers or social media containing information that is not needed. The challenge of data representation is the interpretation of difficult data, for example, in search results data from Google searches or job websites. Matching new data with existing data can be an obstacle to the usefulness of Big Data for central banks.

Using the Irving Fischer Committee on Central Bank Statistics survey data in 2020, Cornelli et al. (2022) show some obstacles facing Asian central banks in implementing Big Data and machine learning techniques. The results of the survey in 2020 results from Asian central banks again revealed similar obstacles as in the previous survey, that is, the reliability of information technology (IT) infrastructure, legal aspects of privacy protection, fair algorithms and data quality. The challenge is that while a reliable IT infrastructure should provide adequate computing and software services, it is expensive and necessitates fending off cyber-attacks. The challenge of a fair algorithm is that neither the programme nor the rules of the algorithm are impartial, for example, processing ethnic and gender data. Asian central banks still face legal challenges to protect privacy and data quality as the results of the BIS 2020 survey show (Cornelli et al., 2022).

3 Big Data Developments in Bank Indonesia

3.1 The Benefits of Big Data in Supporting Policy Formulation

Big Data has several characteristics known as the 5V, i.e., volume (on account of the enormous data size), velocity (due to the high-speed accumulation of data), variety (considering the diverse data types and formats), veracity (meaning the accuracy and reliability of the data), and value (the added value and usefulness of the data). These characteristics make Big Data potentially useful as a new data tool to strengthen the analysis and assessment of policy formulation in central banks. This is also supported by the rapid development of innovations in technology for efficient data collection (e.g., Application Programmable Interface (API)) and fast data processing (e.g., distributed computing) to extract value-added information. That being said, in processing Big Data, the selection of the right methodology is critical for ensuring the accuracy and quality of the resulting data.

The development of Big Data analytics at Bank Indonesia has been ongoing since 2015 and is aimed at supporting the process of policy formulation and supervision in the three main policy sectors, i.e., Monetary/Market, Macroprudential, and Payment System. The Big Data development utilises structured and unstructured data from various sources, both internally from Bank Indonesia, such as payment system data, as well as data from external parties such as news articles and online portals. The applications of Big Data at Bank Indonesia can be classified into four main areas:

a. **Data lags/data gaps**: Big Data fills lags in the compilation of data (data lags) by providing new data/indicators faster or with a higher frequency, so that it can be used as a leading indicator, including in support of nowcasting. For example, processing the residential property price index using sales advertisement data from online property portals available on a monthly basis can be used to complement the results of surveys conducted on a quarterly basis.

b. **Economic and market behaviour and forecasting**: this area includes macroeconomic analysis and prediction as well as the behaviour of market participants and financial systems. For example, daily SBN (Government Securities) transaction data by foreign investors can be combined with various macroeconomic and market indicators to predict foreign investor behaviour (Widjanarti et al., 2021).

c. **Network analysis**: this field analyses the interrelationships between economic actors and the financial system, which can be used for monitoring systemic risk. For example, using BI-RTGS data, the core-periphery model described below can be applied to identify banks that have an important role in the payment system using a network analysis approach (Ari et al., 2018).

d. **Public perception**: this approach identifies public perceptions and expectations of Bank Indonesia policies and involves the use of articles or news to measure the credibility of monetary and macroprudential policies (Wibisono et al., 2021).

3.2 Big Data Development Design/Framework at Bank Indonesia

In general, Big Data development at Bank Indonesia goes through the following steps:

1. **Identify needs**. Big Data development begins with the identification of needs submitted by the departments and offices, complete with the type of data to be generated (e.g., index, network or interrelationships between actors), frequency of data (e.g., daily, monthly) and a clear intended utilisation purpose or plan of the data/indicators to be developed.

2. **Design**. Next, the identification of data sources and methodologies to be used is carried out. The choice of methodology is adjusted to the type of data to be generated, for example:

    a. The sentiment index (public perception) uses ML techniques, especially supervised learning to determine the sentiment of an article (positive, negative or neutral), which is then used to calculate the index,

b. The property price index uses text mining techniques to perform data cleansing, which is then processed to calculate the index.

The data sources and methodologies are included in a Terms of Reference (TOR) document as well as the objectives and completion schedule as a reference for development.

3. **Data collection.** As for data already available at Bank Indonesia, such as payment system transaction data and reporting data, access to such data adheres to the data/information governance policy at Bank Indonesia. Meanwhile, external data will be collected through several mechanisms, for example, Secure File Transfer Protocol (FTPS) for data obtained in collaboration with industry or other institutions.

4. **Processing and analysis.** This stage begins with the pre-processing to prepare the data obtained that will be used in processing or forming ML models. For data in the form of text, for example, the tools include word tokenisation (dividing sentences into words), normalisation (uniforming the form of words into lowercase or uppercase letters) and omitting stop words.

   In case the method used is supervised learning, it may be necessary to annotate and add labels manually to the data that will be used as training data, for example giving positive, negative or neutral labels for sentences that play a role in developing the sentiment model. The annotation process is carried out by employees who master the relevant domain knowledge by referring to the guidelines previously prepared. Furthermore, the development of ML or Big Data models is carried out to generate data or indicators. Testing or validation is undertaken to ensure the required level of accuracy of the model that has been developed, such as calculating Precision and Recall, and comparing the resulting indicator with other data or indicators.

5. **Dissemination**. After the result indicators of ML or Big Data Analytics models have been developed, the resulting indicators are disseminated to users for use in assessment and policy formulation. In utilising such data, users can also evaluate the indicators and can propose further development or model recalibration if there are dynamics in the data source or the need for the indicators that have been produced. The stages of Big Data development are best carried out within the statistical development framework contained in the General Statistics Business Process Model (GSBPM).[1]

3.3 Big Data Projects Developed at Bank Indonesia

Since 2015, Bank Indonesia has produced more than 40 Big Data analytics use cases, covering those that are routinely used to strengthen policy formulation and those that are used to meet the needs of topical assessments. The development of Big Data

1. General Statistical Business Process Model is a framework for describing statistical production in a more comprehensive and process-oriented manner. GSBPM has been adopted by several international institutions as a framework for establishing statistics, e.g., Eurostat, UNStats, etc.

analytics uses various methods of ML, text mining, data mining, network analysis and computer visualisation supported by an infrastructure that has distributed processing capabilities. In addition, the data used is obtained from various data sources, including data from online portals (job advertisements, property sales advertisements, automotive sales advertisements and news portals), Bank Indonesia's payment system infrastructure, industry reporting systems and satellite images.

Some examples of Big Data analytics use cases that have found application in policy formulation are as follows:

a. Monetary Sector/Market

1. Job vacancy index as an indicator of labour demand. This use case utilises job advertisement data which is processed using text mining techniques.
2. Behavioural modelling and the projected flows of foreign investors in the government securities market. This use case utilises various economic and financial indicators as input for a model built with ML algorithms to predict investor behaviour, whether buying or selling (Widjanarti, 2021).

b. Macroprudential Sector and Financial System Surveillance

1. Property price indicators in the secondary market that utilise online property advertising data and are processed using text mining techniques.
2. Network analysis and interconnectedness of banks by utilising data from Bank Indonesia's payment system infrastructure which is processed using core-periphery and ML algorithms. (Ari et al., 2018).

c. Payment System Sector

1. Network analysis and interconnectedness of payment service providers from Bank Indonesia's payment system infrastructure data.

3.4 Bank Indonesia Co-operation with Various Parties to Support Big Data Development

One of the challenges in the development of Big Data analytics is the availability of data sources and the historical continuity of the data. In order to ensure regular data availability, Bank Indonesia collaborates with various parties, i.e., with online property portals, online automotive portals, as well as the largest FinTech and e-commerce players in Indonesia. The co-operation is set forth in a Non-Disclosure Agreement (NDA) between Bank Indonesia and the data providers, which includes the type and structure of the data exchanged, the delivery period, as well as the rights and obligations in order to maintain the confidentiality of data from each party.

With the rapid development of ML algorithms and technology, data scientists and researchers at Bank Indonesia need to continue to be equipped with the latest methodologies and tools that can be used to improve the quality of Big Data analytics. To develop its resource capabilities, Bank Indonesia actively participates in various Big Data workshops and seminars, both nationally and internationally. In addition, Bank Indonesia also conducts in-house training and consultations with academics or experts relevant to Big Data analytics.

3.5 Plan for Future Big Data Development

Bank Indonesia's Big Data development plan in the future will focus on five development areas:

1. **Integration into Data Centre.** In the future, the indicators generated by Big Data analytics will be integrated into the Bank Indonesia Data Centre infrastructure. This will make it easier for users to conduct assessments and analysis, because the indicators can be combined with other data or indicators through one integrated portal.

2. **Utilisation of Self-service Analytics in a Bank Indonesia-Wide way.** In the future, Big Data development will be carried out in a self-service analytical infrastructure that provides flexibility for users to develop or analyse the datasets independently.

3. **Optimising Granular Data or High-Frequency Data for Policy Formulation**. This development will incorporate, for example, Bank Indonesia's Integrated Commercial Bank Reports (LBUT) and BI-FAST data.

4. **Development of Surveillance Technology**. In addition to policy formulation, granular data will also be optimised for surveillance purposes by both banks and non-bank institutions, which could result in an early warning model for the banking sector and behavioural analysis of the interconnectedness of Payment Service Providers (PJP).

5. **Strengthening of Data Governance.** By utilising granular data for Big Data development, it is necessary to strengthen the data governance framework to prevent the risk of data *leakage* (*data breach* and *data leak*) and other data security incidents. This extends to the formulation of data security policies, procedures and regulations in the use of granular data, and to the use of technology and infrastructures that can mitigate risks to data security.

4. Using Big Data to Support Monetary Policy Formulation

4.1 The Benefits of Big Data in Supporting Monetary Policy Formulation

To reach the goal of achieving and maintaining stability of the Indonesian rupiah, Bank has implemented an inflation targeting (IT) monetary policy framework since July 2005. The mandate considers the policy framework and institutional aspects set out by the central bank law. Within this framework, inflation is an overriding policy objective. In the face of changing dynamics and challenges to the economy, Bank Indonesia continues to make various improvements to the monetary policy framework to strengthen its effectiveness. The monetary policy framework is supported by three main pillars, which are (i) dynamic and measurable monetary policy that is ‘ahead-of-the-curve’; (ii) an integrated monetary management with the aim of developing an inclusive and modern money market; and (iii) policy synergy for sustainable economic growth. The implementation of the monetary policy framework is based on several fundamental principles, including that it is forward-looking and based on data and research.

A forward-looking policy direction is essential, considering the time lag of the monetary policy transmission mechanism. In formulating monetary policy, Bank Indonesia uses FPAS (Forecasting and Policy Analysis) to obtain a coherent view of the economic assessment, the main policies that need to be taken and their implications for future research/assessments. This FPAS integrates forecasting and simulation and supports the implementation of IT, which requires a forward-looking policy framework that can anticipate economic conditions pre-emptively, and also because IT involves several aspects (reporting systems, databases, information technology and economic modelling).

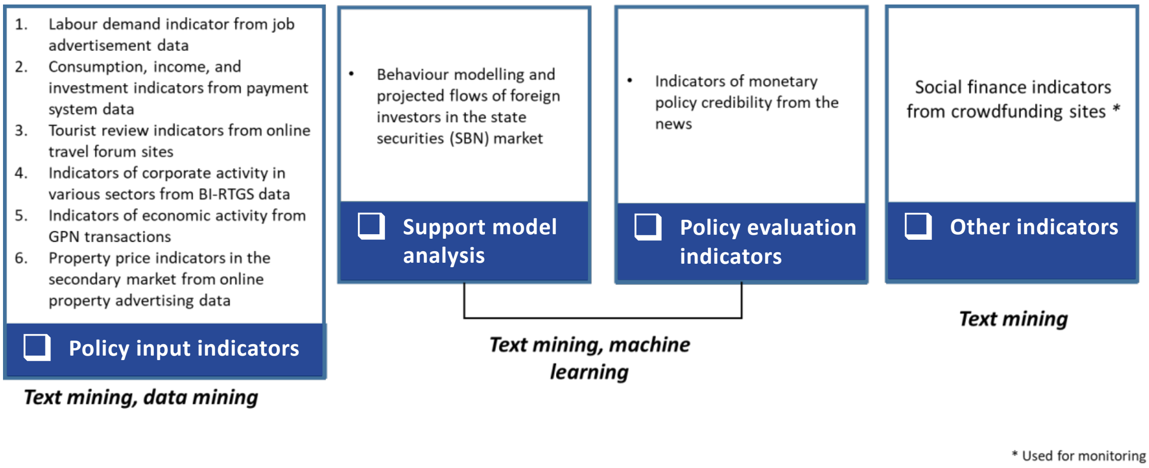
In line with Bank Indonesia’s efforts to develop Big Data analytics, the implementation of FPAS is also supported by using various indicators and models generated from Big Data analytics. The use of Big Data analytics can help overcome data acquisition constraints, especially during the COVID-19 pandemic, as traditional statistical data and existing models are often inadequate to support the decision-making process because they have low data frequency and some of them are only produced with a lag. In particular, the impact of the COVID-19 pandemic changed the behaviour of economic agents rapidly, which required analysis using higher frequency, real-time and granular data.

Currently, the use of Big Data analytics in monetary policy formulation can be broadly divided into several groups: (i) indicator groups for policy formulation inputs; (ii) macroeconomic, monetary and money market analysis/projection models supporting and/or feeding into other models; (iii) indicator groups for policy evaluation; and (iv) other supporting indicator groups. Input indicators and models generated from Big Data analytics support policy formulation. Meanwhile, policy evaluation indicators are used to analyse the impact of policy communication after the formulation of monetary policy. The other supporting indicators are financial indicators from crowdfunding sites specifically formed to support sharia economic assessments.

4.2 Types of Big Data Projects Developed in the Monetary Sector

To date, Bank Indonesia has developed more than six Big Data analytics projects relevant to the monetary sector. More specifically, five projects are for determining policy-supporting indicators; two are Big Data model development projects, i.e., behavioural modelling and projections of foreign investor flows in the SBN (Government Securities) market; and one is a monetary policy credibility indicator project (Figure 4.1).

Figure 4.1
Big Data Indicators in the Monetary Sector



Similar to the use of econometric models to develop indicators and conduct projections, Big Data analytics are not that different from conventional econometric methods. The analysis uses several well-established methodologies, which usually use contemporary data sets that can be complex in terms of size and dimension. But, depending on existing information, the need for additional information from various sources is not always easy and often requires a particular set of capabilities and access. In the context of AI and Big Data analytic techniques, these would include text mining, machine learning, text agent-based modelling, network analysis, statistical methodologies and other modelling techniques (Kurniati, 2017). Accordingly, most Big Data Analytics projects use text mining, data mining and ML.

Meanwhile, the Big Data analytics revolution in payment system data processing offers richer data sources and techniques for understanding and analysing financial system stability. To that end, Bank Indonesia has also developed various indicators as input for policy formulation that utilise payment system data (BI RTGS) to monitor economic activity developments in high-frequency data. The indicators that have been developed from payment system data serve as indicators for consumption, income and investment. In addition, indicators for the corporate sector were also developed using Big Data analytics from BI-RTGS transaction data. This indicator has a good relationship with various sectors' gross domestic product (GDP), so it can be used to monitor economic activities (Sulistiawati, 2021).

In order to map economic conditions, mainly supply and demand, Bank Indonesia also utilises several indicators generated from Big Data analytics. The indicators are labour demand indicators formed from job advertisement data. In addition, Bank Indonesia also generates tourist review indicators from online travel forum sites. The indicators are extracted from various reviews on TripAdvisor, which are then used to identify the advantages and disadvantages of tourist locations so that the advantages of tourist destinations in Indonesia can be compared to destinations in other countries (Jabbar, 2021).

4.3 Examples of Big Data Products to Support Monetary Policy Formulation

This section presents several examples of Big Data analytics that have been developed and utilised to support the formulation of monetary policy at Bank Indonesia.

1. **Job Vacancy Indicator**. The job vacancy index serves as a labour demand indicator and consists of nine sectors for all job vacancies. This index is available monthly and on an earlier basis (with a one-week lag). The data used are sourced from various print media in Indonesia and a job portal in collaboration with Bank Indonesia.

   The indicator is created by using text mining and data mining to process job advertisement text files (e.g., pre-processing, duplicate removal and information extraction). This indicator has been regularly used to support monthly macroeconomic assessments, particularly household consumption. The evaluation results show that the job vacancy indicator has a good relationship (or correlation) with the income index, so it can be used as a proxy indicator for supporting household income assessments, which are the determinants of household consumption (Figures 4.2 and 4.3).

Figure 4.2
Job Vacancy

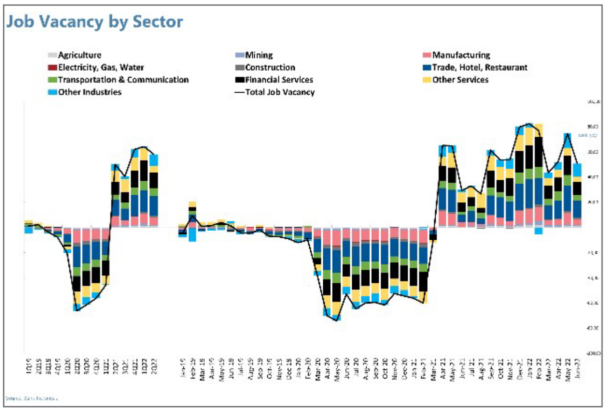
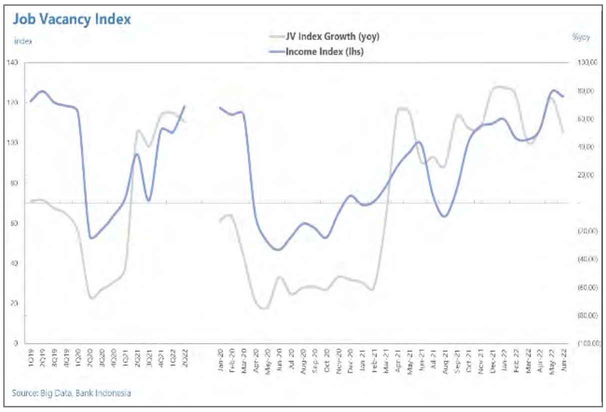


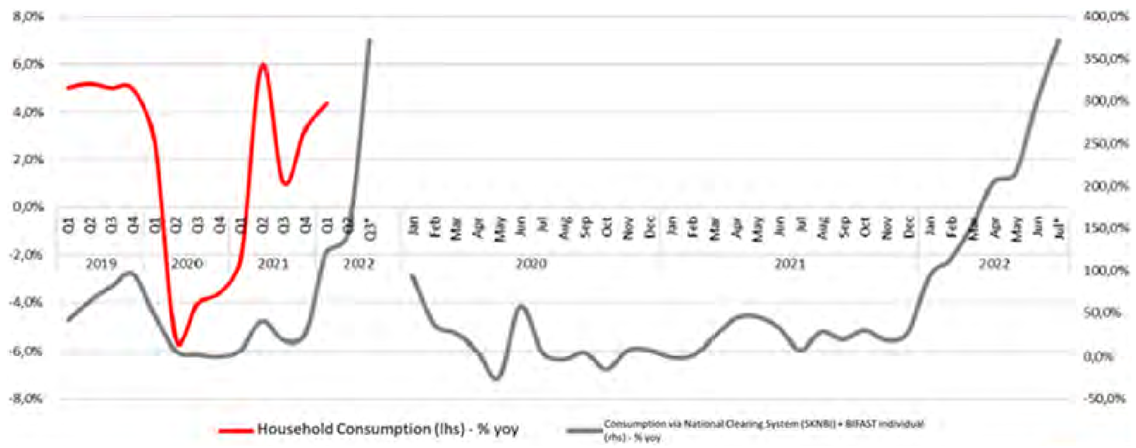
Figure 4.3
Job Vacancy and Income Index



2. **Household Consumption Indicator**. This indicator was constructed as a high-frequency indicator for the early detection of developments in household consumption. As such, it reflects the major role of household consumption in Indonesia's GDP. The early monitoring of developments in household consumption will provide an important input in determining the direction of the latest economic developments. This indicator also bridges the availability of household consumption indicators in GDP, which are published by Statistics Indonesia (BSP) every quarter with a publication lag of one month. By applying Big Data analytics techniques to granular payment system data, Bank Indonesia can generate faster indicators of household consumption, which can be accessed a few days after the end of the reference period. With its strong correlation with economic growth data, this index can serve as a first-pass proxy for household spending in Indonesia (Zulen et al., 2021). This indicator is created using Bank Indonesia's National Clearing System (SKNBI) and BI-FAST data (since June 2022). The data is processed by text mining using a rule-based model to categorise the purpose of public transactions using SKNBI and BI-FAST. Due to indications of a shift of consumers

from using SKNBI to BI-FAST, the utilisation of BI-FAST data only started from June 2022, in line with the implementation of BI-FAST starting in December 2021. So far, this household consumption indicator has a reasonably good correlation with household consumption published by BSP, especially during the pandemic period (Figure 4.4).

Figure 4.4
Household Consumption Indicator + BI-FAST and Household Consumption



3. **Behaviour Modelling and Projected Flows of Foreign Investors in the SBN (Government Securities) market.** Investors can be classified as real money or traders based on their portfolio management behaviour. Mapping foreign investors will help identify the dominant group of foreign investors sensitive to financial market sentiment so that the central bank can formulate a more appropriate policy response (Widjanarti et al., 2021). In line with that, Bank Indonesia initiated behaviour modelling. It projects flows of foreign investors in the SBN (Government Securities) market by utilising Big Data analytics to explain and predict foreign investors' behaviour and capital flows. The model is prepared with individual investor behaviour analysis, which includes decision tree algorithms on market and fundamental variables and LIME (Local Interpretable Model-Agnostic Explanation). The results obtained from Big Data analytics include (i) the mapping of individual investor behaviour and classification, (ii) foreign investor decisions, and (iii) capital flows projection. In line with the increasing number and dynamics of foreign investor behaviour in the government securities market, the model needs to be re-calibrated for investor classification with an enhanced methodology through data-driven Big Data analytics techniques. In addition, Big Data Analytics also has good potential in predicting the behaviour of individual foreign investors in various scenarios of economic indicators or financial markets.

4. **Indicators of Monetary Policy Credibility.** The aim of this indicator using Big Data analytics, based on news in the mass media, is to measure public perceptions of the credibility of Bank Indonesia's monetary policy. The measured credibility aspects are aligned with the existing Bank Indonesia Policy Survey, i.e., formulation, co-ordination, communication and effectiveness. This index also replaces the previous measurement of monetary policy credibility, measured through a survey of Bank Indonesia stakeholders. The data is sourced from news articles (cyber library). These indicators can be found in Figure 4.5.

The method used is text mining with ML to classify perceptions of the credibility aspect of Bank Indonesia. This ML model was built using training data from news sentences. In line with the developments in monetary policy, the formation of this index also underwent adjustments. Recently, for example, the adjustment is made to capture the unprecedented policies adopted by Bank Indonesia during the COVID-19 pandemic. With such an adjustment, the credibility index captures public perceptions of the unconventional monetary policy adopted by Bank Indonesia during the COVID-19 pandemic. Going forward, in line with the normalisation of Bank Indonesia's monetary policy, the establishment of the credibility index also needs to be readjusted.

5. Using Big Data to Support Macroprudential Policy Formulation

5.1 The Benefits of Big Data in Supporting Macroprudential Policy Formulation

In implementing macroprudential policy, Bank Indonesia operates with a policy framework underlying the formulation of macroprudential policies to strengthen the implementation and effectiveness of policies, including through the implementation of comprehensive surveillance. In 2020, Bank Indonesia revisited its macroprudential policy and surveillance framework, which was later called the Dynamic Integrated Macroprudential Policy and Surveillance Framework (DIMPS). DIMPS is a dynamic macroprudential policy and surveillance framework that is integrated within macro-financial (time varying) and micro-surveillance (cross section) dimensions. The framework is supported by economic and financial inclusion and is used to maintain financial system stability through Bank Indonesia's policy mix. DIMPS is implemented through an operational strategy that consists of the following stages (Agung et al. 2021), which are reproduced in Figure 5.1:

1. **Systemic Risk Identification, Financing and Inclusion** refers to the monitoring phase of the financial system vulnerability indicators and other indicators that may affect the financial system or become sources of shock (such as exchange rates and capital flows indicators) This includes indicators derived from the use of technology and digital innovation, for instance, the implementation of Big Data and ML.

2. **Systemic Risk Assessment, Financing and Inclusion** concerns the measurement phase of the potential impact arising from the increased risk that has been identified in the previous phase, which includes forward-looking assessments, notably projections, policy simulations and stress test analysis.

3. **Policy Recommendations and Implementation** is the development phase of the design and formula of policy instruments, formulation of regulations, co-ordination mechanisms in policy implementation, both internal and external co-ordination with other authorities, as well as policy communication.

4. **Policy Evaluation and Monitoring** is the phase of monitoring the implementation and evaluation of policy effectiveness, including thematic examinations if necessary.

Figure 5.1
DIMPS Cycles

Source: Bank Indonesia.

Observing the immense and rapid development of the latest technology, including in the financial sector, it is undeniable that the current trend of accelerating data creation through real-time data formation, with its associated large volume and high diversity, will continue to increase. Ultimately, this may have a positive and significant impact, particularly for central banks that are still facing challenges in conducting assessments or analyses, due mainly to the lack of available data (including the problem of data publication lags). Concerning macroprudential policy formulation, Bank Indonesia has utilised Big Data obtained, among others, from Real-Time Gross Settlement (RTGS) data, Bank Indonesia's National Clearing System (SKNBI), Financial Information Service System (SLIK) and other external data sources, to support the stages in the process of formulating macroprudential policies namely:

1. Risk identification through monitoring or supervision of indicators in the financial system which is carried out regularly or if deemed necessary (topical);

2. Strengthening assessments in the financial system, both routine and topical assessments; and,

3. Strengthening the evaluation and monitoring of policies.

Thus, the monitoring and assessment process is expected to capture behaviour and phenomena in the financial sector in a timelier manner. Moreover, a huge amount of granular information, extracted from extremely large data sets, increases the possibility of gaining more comprehensive results and more valid and reliable forward-looking indicators as well.

For instance, in the risk identification or monitoring phase, the identification of structures and interconnectedness relevant to transactional aspects carried out by banks is conducted using network analysis and ML methods. In addition, data exploration has been carried out using ML to analyse vulnerabilities in the household sector and the quality of debtors or bank credit as part of the financial system assessment. Looking ahead, the potential for the development and exploration of Big Data use is still immense in the recommendation and policy implementation phase.

5.2 Types of Big Data Projects Developed in the Macroprudential Sector

In support of the formulation of macroprudential policies, several indicators have been developed with the help of Big Data. Since 2015, more than ten types of Big Data projects have been applied to support the formulation of macroprudential policies. In general, such indicators can be classified into three groups based on their purposes, namely risk identification/monitoring indicators, systemic risk assessment models/tools and policy evaluation indicators (Table 5.1).

Table 5.1
Big Data Projects/Products Developed to Support Macroprudential Policy Formulation

RISK IDENTIFICATION	RISK ASSESSMENT	POLICY EVALUATION AND MONITORING
Labour demand indicator from job advertisement data	Credit risk nowcasting	Indicators of macroprudential policy credibility from the news
Consumption and income indicators from payment system data	Job vacancy indicators based on labour classification	
Property price indicators in the secondary market from online property advertising data	Characteristics of debtors as predictors of individual mortgage default risk	
Banking network analysis and interconnectedness from BI-RTGS and PUAB data		
Banking network analysis and interconnectedness from foreign exchange transaction data (BI-RTGS)		
Sectoral corporate activity indicators from BI-RTGS data		
Labour vulnerability indicators from news		
Aggregator of news or issues related to the financial system		

Most of the Big Data products that have been utilised were developed using ML and text mining methods. For example, the identification of debtor characteristics that affect the probability of default (PoD) of Individual Mortgage or Home Ownership Loans (KPR) is obtained from ML techniques. Furthermore, ML is also applied to assess banking sector vulnerabilities, such as credit risk nowcasting which is processed from SLIK data. In practice, some of the assessments that utilise indicators from Big Data have been carried out thematically rather than regularly, considering the complexity of the methodology and the time required to produce an assessment or analysis using Big Data. Meanwhile, such indicators are used to support the main indicators in routine assessments carried out monthly to strengthen policy formulation discussed in the Board of Governors' Meeting (RDG). But several indicators also serve as assessment tools and are monitored regularly, including those obtained through network analysis and interconnectedness of the BI-RTGS data.

As mentioned above, one of many methods to monitor and supervise the financial system is through network analysis and interconnectedness of the BI-RTGS data on a weekly basis. This is conducted to identify the structures and interconnections related to transactional aspects carried out by banks as one of the aspects related to systemic risk assessment. Regarding policy evaluation and monitoring, there are indicators of macroprudential policy credibility generated from news in the public media using the text mining method. In this case, the public's perception of the credibility of macroprudential policies is measured for evaluation purposes on a quarterly basis.

5.3 Examples of Big Data Products to Support Macroprudential Policy Formulation

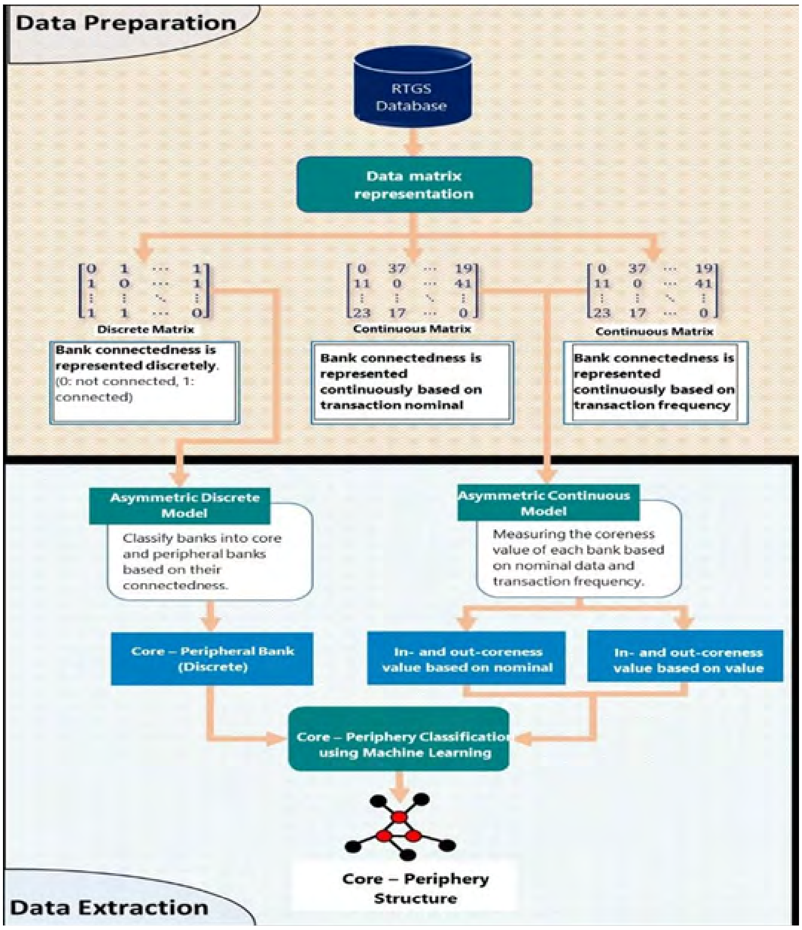
Of the various indicators derived from the use of Big Data to support the formulation of macroprudential policies, some use case samples will be elaborated on in more detail. They represent Big Data applications in the stages of macroprudential policy formulation, particularly for monitoring or risk identification purposes, risk assessments and policy evaluation and monitoring.

1. **Monitoring or Risk Identification Using Network Analysis and Banking Interconnectedness from BI-RTGS Data (Ari et al., 2018)**

   Given the importance of the large value payment system (BI-RTGS) and its relation to financial system stability and systemic risk, it is imperative for the central bank to analyse the structure of the interrelationships of each participant in the payment system. Therefore, it is necessary to identify the structure of linkages between actors (banks) in the BI-RTGS payment system using network analysis methods, which can be used to complete the analysis of the systemic impact on the financial system as well as identify banks that are included in the core bank category, i.e., banks that have an important role in the BI-RTGS payment system. The data used is structured data sourced from BI-RTGS, both in nominal and transaction volumes from four types of transactions: (1) transactions between participants; (2) transactions for customer needs; (3) Interbank Money Market transactions; and (4) foreign exchange transactions.

Conceptually, the network structure in the interbank market is tiered where there is a group of densely interconnected banks that also act as intermediaries between other banks that are sparsely connected. This network is divided into two groups, i.e., the core group which is a collection of nodes on the network that are densely connected, and the periphery group which consists of nodes that are sparsely connected. In the BI-RTGS payment system, the representation of a network can be made by defining each bank as a node and interbank transaction data as an edge (link). The network observed in the payment system is a directional network, and the appropriate models are the asymmetric discrete model (Craig and von Peter, 2010) and the asymmetric continuous model (Boyd et al., 2010). An asymmetric discrete model is used to classify core and periphery banks based on inter-bank connectivity so that core-periphery partitions can be generated. The asymmetric continuous model measures how strong a node can act as a core through a "coreness" measure based on nominal or transaction frequency. Furthermore, ML methods are used to classify core banks and periphery banks based on input data in the form of discrete partition results (banks have been labelled as core or periphery) and out-coreness and in-coreness values as characteristics of each bank. Machine learning will pick up which banks are labelled as core (discrete) and have similar characteristics to produce a classification model using the SVM algorithm. The core-periphery bank classification framework is illustrated in Figure 5.2.

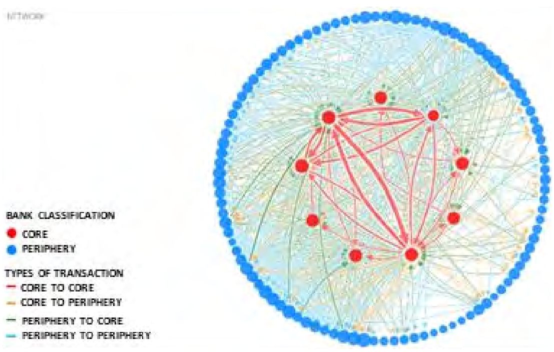
Figure 5.2
Core-Periphery Bank Classification Framework



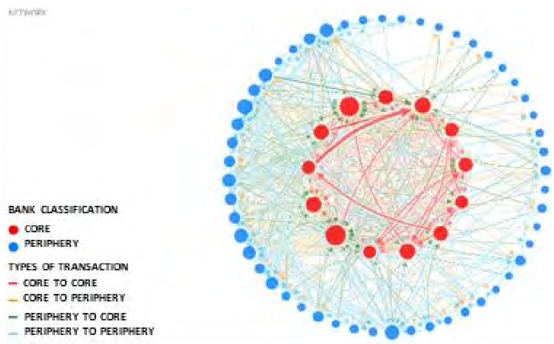
The results of the analysis based on transaction data show that almost all BI-RTGS participating banks actively transact with each other for the four types of transactions. By contrast, the average transaction shows that the network density is relatively moderate. This indicates that the transaction flow tends to be centred on a few banks (interbank tiering). In addition, from the partition results obtained, it can be inferred that the core-periphery model is compatible with the BI-RTGS network for all types of transactions.

The discrete model shows that several banks can be identified as core banks. Results of the continuous model are generally in line with the results of the discrete model because the banks with the highest coreness values are also banks that always appear as core in the discrete model. The SVM classification further indicates that there are fewer banks indicated as core based on interbank connectivity, transaction nominal data and transaction volume data. From the classification using SVM, a visualisation of the core-periphery structure for each type of transaction can be obtained, as shown in Figure 5.3.

Figure 5.3
Illustration of Core-Periphery Structure Visualisation



Total RTGS Core-Periphery



Forex RTGS Core-Periphery

2. **Strengthening of Risk Assessment Using Determinant Analysis of Debtor Characteristics and Prediction Models of Individual Mortgage Default Risk (Surjaningsih et al., 2022)**

The assessment of the household sector is part of the financial system assessment regularly conducted within the macroprudential policy framework. The probability of default (PoD) of a mortgage as measured through a bank's non-performing loans (NPL) is perceived as a potential source of shock to the financial system. Hence, this assessment aims to investigate changes in Indonesian households' financial resilience, specifically the PoD indicator, before and during the pandemic by utilising an individual level data set, i.e., the SLIK, which provides comprehensive mortgage data. With the high diversity and complexity of debtor information, the limitations of statistical methods are overcome by ML methods to capture complex nonlinear relationships.

The processed data is sourced from the Individual Debtor Mortgage Report managed through SLIK with an active credit facility period from December 2019 to June 2021. The dataset is classified into three periods, i.e., the pre-pandemic period (dataset 1), the peak of the pandemic period (dataset 2), and the recovery period (dataset 3). This assessment uses a ML approach to produce the best model for studying individual characteristics to predict household credit vulnerability as conducted by Teply and Polena (2020) and Song et al. (2020). The study finds that the XGB algorithm is appropriate for modelling a default event.

From this assessment, it can be concluded that the most significant determinant of the PoD for the pre-pandemic period is the project location, which implies that the PoD tends to be high in particular regions during the pre-pandemic period. Other important features are the remaining maturity and the maturity (Table 5.2).

Table 5.2
Importance of Features

	Dataset 1	Dataset 2	Dataset 3
1	project location	interest rate	interest rate
2	remaining maturity	current instalment value	current instalment value
3	maturity	project location	frequency of restructuring
4	frequency of restructuring	maturity	remaining maturity
5	interest rate	remaining maturity	maturity
6	current instalment value	initial instalment value	LTV ratio
7	initial instalment value	frequency of restructuring	project location
8	age group of borrower	gross income	age group of borrower
9	loan ceiling	loan ceiling	field of occupation
10	gross income	age group of borrower	initial instalment value
11	field of occupation	contract type	contract type
12		field of occupation	gross income
13		sequence of credit facility	loan ceiling
14			sequence of credit facility
15			property type

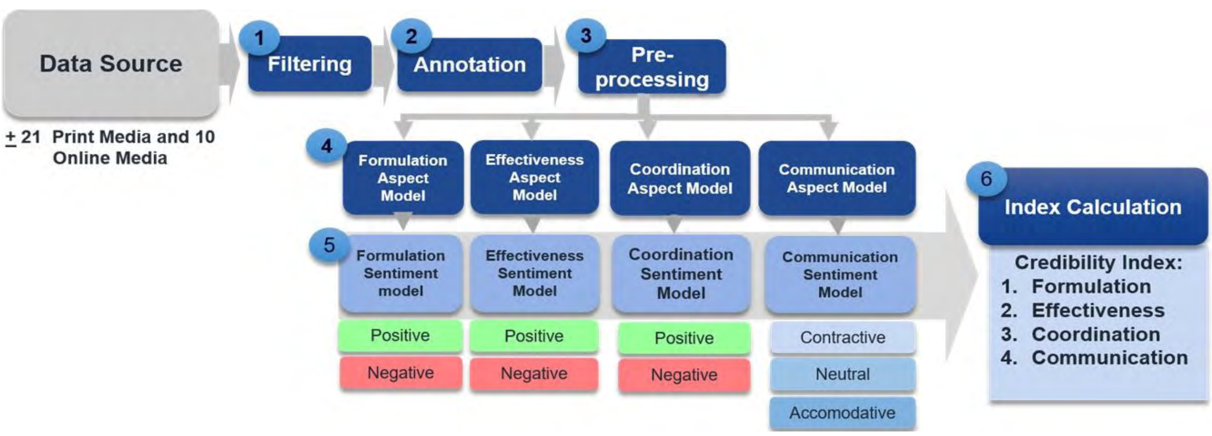
This assessment also indicates that there are differences in the pattern of characteristics of mortgage debtors that affect PoD before the pandemic, during the pandemic and during the recovery period. In the pre-pandemic period, the project location was the most important feature. In the period of the highest NPL during the pandemic, there was a shift in the pattern, especially the interest rate and current instalment features, which became increasingly important in the default prediction model. In the recovery period, interest rates became the most significant feature in the default event prediction model. These changes in importance between the three datasets indicate that during the pandemic period, the interest rate (cost of debts) and current instalment value became increasingly significant to the debtors' repayment capability.

3. Strengthening of Policy Evaluation through the Use of Macroprudential Policy Credibility Indicators from News (Jabbar et al., 2022)

Measuring policy credibility *via* periodic public surveys has several weaknesses, including the potential for differences in perceptions in understanding the questions in the survey and the inappropriate selection of respondents by third parties conducting the survey. On the other hand, the development of digital news, which is currently one of the sources of public information, opens up opportunities for other sources in measuring the credibility of macroprudential policies. Therefore, it is necessary to develop an alternative measurement of credibility through the collection of public information, for instance, news data from various media (print/online), using Big Data technology. The preparation of alternative credibility measurements is carried out to measure public perceptions of the credibility of Bank Indonesia's macroprudential policies, which are sourced from reports in the mass media. There are four aspects of credibility that are measured, i.e., formulation, effectiveness, co-ordination and communication (conformity between forward guidance and public expectations of macroprudential policy directions). The resulting perception index is expected to provide comparisons between periods, including when pressure occurs or before and after the policy is implemented.

The news sources used are in the form of news articles (text) from the BI Cyber Library. This data is available daily and includes economic and financial news relevant for analysing macroprudential policy responses. News is filtered based on the emergence of keywords pertaining to macroprudential policies and then processed into sentences (sentence splitting). The data period used is from 2013 to 2021. Credibility measurement is carried out by utilising Big Data analytics. More specifically, a ML model is built based on news sentence annotations (credibility intonation labelling). The methodology for measuring policy credibility using news data and text mining in this study is a further development of the study conducted by Zulen et al. (2020). In general, the stages of measuring the credibility index of macroprudential policies from news using Big Data Analytics include filtering, annotation, pre-processing, modelling with machine learning, and index calculations as illustrated in Figure 5.4.

Figure 5.4
Macroprudential Policy Credibility Index Model Workflow



In terms of index calculation, the credibility index is the unweighted average of each of the four measured credibility aspects:

1. formulation: perception of Bank Indonesia's macroprudential policy formulation;
2. effectiveness: perception of the effectiveness of Bank Indonesia's macroprudential policies according to their objectives;
3. co-ordination: perception of Bank Indonesia's co-ordination with the Government and other relevant agencies; and
4. communication: perceptions of expectations of future macroprudential policies disclosed by the public in the news, compared to forward guidance in Bank Indonesia's communication outlets (including, among others, press releases).

The resulting credibility index has a range between zero to 100 per cent. The more news with positive credibility, the closer the index value will be to 100 per cent and *vice versa* This index is updated quarterly for evaluation purposes. The results of the credibility index of Bank Indonesia's macroprudential policy using Big Data technique is in line with the results of the survey on the credibility of Bank Indonesia's macroprudential policy in the period where the same credibility aspects were measured.

6. Big Data Challenges and Areas for Development

6.1 Challenges of Using Big Data

As a supporting tool for the process of formulating macroprudential policies, the use of Big Data cannot be separated from several challenges that often become obstacles. Broadly speaking, such constraints can be grouped into four types, i.e., technical or operational aspects (technicalities), human resources aspects (human resources), product quality aspects (quality assurance) and legal aspects.

The first aspect is related to technical or operational areas, starting with the fact that data of considerable size often presents obstacles in the data cleansing process. The latter is necessary to ensure data quality. In general, large data sets do not contain clean data that is ready to be processed as is. It is possible that there are parts of the data that are not actually needed in the analysis or are even inaccurate. As such data cleansing is necessary, a task requiring extra efforts in practice and can be time-consuming, especially when dealing with large volumes of data. The quality of the data provided by a data source and the limited scope sometimes become obstacles in conducting an analysis. In addition, Big Data processing requires the support of technological equipment/infrastructure with the required capacity and specifications.

Another operational aspect that poses a challenge is the limited sources as well as the continuity of data that can be accessed as input for Big Data analysis. One of the keys to doing Big Data analytics is, of course, data availability. Access to data, both old and new, can be an obstacle, especially with regard to old data that is stored in different and varied formats, often in physical forms. Access to new data also requires effort because permits and licenses are required to legally access non-public data. Both service and information providers who implement Big Data analytics in Indonesia have reported that the data collection phase presents the main challenge. While data in Indonesia is available and plentiful, the sources are scattered, and as such, a lot of effort is needed to obtain integrated data nationally. Therefore, the principle of sharing data and even having more open data is needed.

In line with the principle of open data, the data available in government institutions, including local governments, needs to be presented in a certain standard/platform format, i.e., a format that can be easily re-used and machine-readable as well as interoperable. Thus, open data can increase the use of data sourced from the government, increase government transparency and accountability, and increase community participation in overseeing data development. In addition, the standardisation of government data can be followed by data integration to prevent redundancy.

In addition, and this applies particularly to Indonesia, sources of data and information relevant to the real sector that are granular in nature are still very limited in their availability and are not timely. For example, in the monetary economy, granular data for the assessment of economic growth, both in terms of expenditure and (sectoral) production, is very scanty. Likewise, information appropriate to real sector activities is also still very minimal. In addition, from an operational perspective, Big Data development requires time to deepen needs, data sources, and methodologies, as well as model development, including model tuning in the event that there are adjustments relevant to the dynamics of business needs or data sources.

The second aspect relates to the operators and staff who perform analysis or processing of Big Data. Ideally, the utilisation of Big Data analytics requires operators who are experts in the field of data analysis (data scientists), have analytical skills, computer programming skills, and creativity to determine new methods that can be used to collect, interpret and analyse data. Although data scientists may not develop their own analytic tools, they must be able to sort out the various tools to be used, as well as select and organise the data to be analysed. In case of further needed development, operators must be sufficiently capable and skilled to manage it. Sometimes, there can even be a team of experts who in the end prepares plans for implementing policies or programmes that have been identified. In addition, mastery of Big Data analytics is not a common skill, so that sometimes external consultants are still required. In addition, in terms of numbers, the availability of human resources with special expertise in Big Data analytics is also still very limited.

The third aspect is the issue of product quality (quality assurance). In general, the use of Big Data has a positive impact on accuracy and time efficiency in conducting assessments/analyses. But a precautionary principle is needed for the use of Big Data in policy formulation to ensure the credibility and quality of the policies adopted. In addition, one of the issues that needs to be considered in the use of Big Data is whether the qualitative results of the analysis are in accordance with economic theory or other factors needed in policy formulation. The last aspect relates to the legal requirements for the use of personal and confidential data by the central bank.

6.2 Areas for Further Development

The use of Big Data for the assessment and formulation of monetary and macroprudential policy at Bank Indonesia has the potential to be developed further so that the majority of the resulting products not only support indicators in the assessment process, but become key indicators that are routinely monitored. This is aligned with the development of technology that allows the provision of large and more comprehensive data in almost real time which, with proper utilisation, can support the policy formulation process even further. To that end, there are still several areas for the development of Big Data analytics going forward.

First, as far as operational aspects are concerned, it is necessary to expand co-operation with data owners/providers to increase the coverage and continuity aspects of data sources used for Big Data analytics. In the macroprudential sector, efforts can be made to improve the quality of the data from data sources managed by financial authorities, so that the assessments or monitoring are more reliable, and these can include more granular analytics according to the needs and objectives of the assessment. In other words, there is still a need to improve the quality and coverage of data, especially data that is more real-time, and the availability of an infrastructure for the development of Big Data usage.

In addition, co-operation with external institutions or private sector banks to obtain more granular indicators to enrich existing data sources also has the potential to become an area of development, while considering the confidentiality aspect of individual data. Furthermore, Big Data analytics can also be expanded to include forward-looking analysis. Sharpening the use case for data analytics using granular data is also needed to strengthen the decision-making process. In addition, the development of methodologies to ensure the validity and robustness of the resulting Big Data is also an important development area.

Second, to overcome the challenges of limited human resources, efforts that can be made in the future include improving the quality of human resources, either by recruiting experts (data scientists) or by improving the quality and skills of existing personnel. In addition, strengthening and expanding co-operation of the central bank with other relevant institutions can also be undertaken to close the gap in human resource needs through technical assistance programmes, exchanges of views and (short-term) assignments to other institutions.

Third, in terms of the legal requirements connected with the use of personal and confidential data by the central bank, it may become necessary in the future to take steps to strengthen existing regulations to serve as a legal umbrella for the use of such data by the central bank.

7. Conclusion and Policy Implications

7.1 Conclusion

1. Big Data is increasingly used by central banks in their research, policy formulation and decision-making. Many central banks have reported using Big Data in their operations. In a 2020 survey, more than 40 per cent of organisations reported that Big Data guided decision-making. The use of Big Data in the conduct of policy analysis and central bank decision-making has made great strides in recent years. In particular, the granular data that central banks collect has become an essential knowledge source for decision-makers.

2. Bank Indonesia has developed Big Data analytics since 2015 in support of its formulation of monetary and macroprudential policies. On the monetary side, Bank Indonesia has developed more than twelve Big Data analytics pilot projects in the form of indicators for monitoring economic activity and models to support monetary policy formulation, including money market analysis and policy evaluation. On the macroprudential side, Bank Indonesia has developed more than 20 Big Data projects to support the formulation of macroprudential policies. In general, the indicators can be classified into three categories based on their intended use, i.e., risk identification/monitoring indicators, systemic risk assessment models/tools and policy evaluation indicators.

3. In support of monetary and macroprudential policy formulation, the use of Big Data at Bank Indonesia cannot be separated from several challenges that often become obstacles. Such constraints can be categorised into four aspects, i.e., technical or operational aspects (technicalities), human resources aspects (human resources), product quality aspects (quality assurance) and legal aspects.

    a. **Technical or operational aspects**. The data cleansing process requires extra efforts and can be extremely time consuming, especially when dealing with large volume data. The quality of the data, or lack of thereof, provided by a data source and the limited scope sometimes become obstacles when conducting an analysis. Big Data processing thus requires the support of technological equipment/infrastructure with sufficient capacity and specifications. Another operational aspect is the limited sources and continuity of data that can be accessed as input for Big Data Analytics.

    b. **Human resource aspects**. Big Data analytics requires human resources who are experts in data analysis (data scientists), having the analytical, computer programming skills and creativity to determine new methods that can be used to collect, interpret and analyse data.

    c. **Aspects of product quality (quality assurance)**. The use of Big Data positively impacts accuracy and time efficiency in conducting assessments/analyses. But caution is required in using Big Data in policy formulation to ensure the credibility and quality of the policies adopted. In addition, one of the issues that needs to be considered in using Big Data is whether the quality of the results of the analysis are in line with economic theory and other factors needed in policy formulation.

d. **Legal Aspects**. The use of Big Data requires strengthening the legal base for the use of private and confidential data by central banks.

7.2 Policy Implications

Big Data Analytics is an increasingly vital tool in the rapidly evolving digital era in which we live. The COVID-19 pandemic, in particular, has also changed the behaviour of economic agents and the use of Big Data analytics at central banks for policy making is becoming even more important for improving the quality of statistics, forecasting and nowcasting, and monitoring/surveillance. A central bank's knowledge domain is needed in the development of Big Data analytics. The combination of Big Data technology and central banking can produce a much-improved analysis of policy making. In the future, Big Data analytics will require development in the following areas to overcome the various challenges that are still being faced today:

a. **Operational Aspects**. It is important to expand co-operation with data owners/providers to increase the scope and continuity of data sources used for Big Data analytics.

b. **Human Resources Aspects**. The quality of the available human resources can be improved either through the recruitment of experts (data scientists) or increasing the quality and skills of existing personnel. In addition, improvements in human resources can also be made by strengthening and expanding co-operation of the central bank with other relevant institutions to close the gap in human resource needs through technical assistance programmes and (short-term) assignments to other institutions.

c. **Legal Aspects**. Challenges in this respect can be addressed by formulating provisions to serve as a legal base for the use of personal and confidential data by the central bank.

References and Further Reading

Abraham, F., Schmukler, S. L. and Tessada, J. (2019). Using *Big Data* to expand financial services: Benefits and risks. *World Bank Research and Policy Briefs*, (143463).

Agung, J., Harun C. A. and Derianto, E. (2021). Kebijakan makroprudensial di Indonesia: Konsep, kerangka, dan implementasi. Ed. 1. Depok: Rajawali Pers.

Ari, H., Zikri, M., Kristiawardani, K., Sampe, I. and Hidayat, M. (2018). Identifikasi struktur keterkaitan pelaku dalam sistem pembayaran. LHP Bank Indonesia.

Benchimol, J., Kazinnik, S. and Saadon, Y. (2022). Text mining methodologies with R: An application to central bank texts. *Machine Learning with Applications*, *8*, 100286.

Baghestani, H., and Danila, L. (2014). Interest rate and exchange rate forecasting in the Czech Republic: Do analysts know better than a random walk? *Finance a Uver: Czech Journal of Economics & Finance*, *64*(4).

Bholat, D. (2015). Big Data and central banks. *Big Data & Society*, *2*(1). https://journals.sagepub.com/doi/full/10.1177/2053951715579469.

Boumghar, R. (2019). Machine learning and data science. *at ECB Eurofiling Conference 2019*, Frankfurt.

Boyd, J. P., Fitzgerald, W. J., Mahutga, M. C. and Smith, D. A. (2010). Computing continuous core/periphery structures for social relations data with MINRES/SVD. *Social Networks*, *32*(2), 125-137.

Bruno, G. (2016). Text mining and sentiment extraction in central bank documents. In *2016 International conference on Big Data*, (pp. 1700-1708). IEEE. https://ieeexplore.ieee.org/document/7840784.

Clarida, R., Gali, J. and Gertler, M. (1999). The science of monetary policy: A new Keynesian perspective. *Journal of Economic Literature*, *37*(4), 1661-1707.

Cornelli, G., Doerr, S., Gambacorta, L. and Tissot, B. (2022). Big Data in Asian central banks. *Asian Economic Policy Review*, *17*(2), 255-269. https://www.bis.org/ifc/publ/ifcwork21.pdf.

Craig, B. R. and von Peter, G. (2010). Interbank tiering and money centre banks. BIS Working Papers 322, Bank for International Settlements. https://www.bis.org/publ/work322.pdf.

Cœuré, B. (2017). "Policy analysis with Big Data", Speech at the Conference on Economic and Financial Regulation in the Era of Big Data, organised by the Bank of France, Paris, 24 November.

Delgado-de-la-Garza, L. A., Garza-Rodríguez, G. A., Jacques-Osuna, D. A., Múgica-Lara, A. and Carrasco, C. A. (2021). Does the use of a Big Data variable improve monetary policy estimates? Evidence from Mexico. *Economics and Business Letters*, *10*(4), 383–393. https://reunido.uniovi.es/index.php/EBL/article/view/15253/14490.

Doerr, B. S., Gambacorta, L. and Serena, J. M. (2021a). How do central banks use Big Data and machine learning? *SUERF Policy Briefs No 67*. https://www.suerf.org/docx/f_b0573260a878535507443fc3e6f1e34f_22565_suerf.pdf.

Doerr, S., Gambacorta, L. and Garralda, J. M. S. (2021b). Big Data and machine learning in central banking. BIS Working Papers, (930). https://www.bis.org/publ/work930.pdf.

Doko, F., Kalajdziski, S. and Mishkovski, I. (2021). Credit risk model based on central bank credit registry data. *Journal of Risk and Financial Management*, *14*(3), 138. https://www.mdpi.com/1911-8074/14/3/138.

European Commission (2016), *Big Data analytics for policy making*, Directorate-General for Informatics. https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf.

European Central Bank (ECB). (2017). *MMSR–IT Appendix for reporting agent-version 3.0*. European Central Bank. https://www.ecb.europa.eu/stats/money/mmss/shared/files/MMSR-IT_appendix_for_reporting_agents.pdf.

ECB. (2019). *Big Data and machine learning initiatives at the ECB*. Bank of Italy and BIS Workshop on Computing Platforms for Big Data and Machine Learning, Rome, 15 January 2019.

Eskici, H. B. and Koçak, N. A. (2018). A text mining application on monthly price developments reports. *Central Bank Review*, *18*(2), 51-60.

Galar, M., Fernández, A., Barrenechea, E., Sola, H. B. and Herrera, F. (2012). A review on ensembles for the class imbalance problem: bagging, boosting and hybrid-based approaches. *IEEE Transactions on Systems Man and Cybernetics Part C, 42*(4), 463-484.

Haryono, E. (2022). *Data expansion to strengthen policymaking*, 15 February. https://www.bi.go.id/en/publikasi/ruang-media/news-release/Pages/sp_245122.aspx.

Hassani, H., Huang, X. and Silva, E. (2018). Digitalisation and Big Data mining in banking. *Big Data and Cognitive Computing*, *2*(3), 18. https://www.mdpi.com/2504-2289/2/3/18.

Israël, J-M. (2017, March). AnaCredit and RIAD. *BIS-BI-ECB Regional Seminar 2017*.

Israël, J-M., Damia, V., Bonci, R. and Watfe, G. (2017). The analytical credit dataset: A magnifying glass for analysing credit in the euro area. *ECB Occasional Paper 187*. https://www.ecb.europa.eu/pub/pdf/scpops/ecb.op187.en.pdf.

Israël, J. M. and Tissot, B. (2021). Incorporating microdata into macro policy decision-making. *Journal of Digital Banking*, *6*(3), 233-250.

Jabbar, M., Widjanarti, A., Zulen, A. A. and Okfantia, A. (2021). Sentiment analysis of tourist reviews from online travel forums for improving the Indonesian tourism sector.

Jabbar, M. A., Wibisono, O., Praptono, N. H. and Zulen, A. A. (2022). Measuring macroprudential policy credibility using machine learning. *Forthcoming BIS Paper.*

Kurniati, Y. (2017). *Opening remarks. IFC-Bank Indonesia Satellite Seminar on "Big Data",* Bali, 21 March. https://www.bis.org/ifc/events/big_data_mar2017/bigdata_welcome1.pdf.

Lo, I. S., McKercher, B., Lo, A., Cheung, C. and Law, R. (2011). Tourism and online photography. *Tourism Management*, *32*(4), 725-731.

Lucca, D. O. and Trebbi, F. (2009). *Measuring central bank communication: An automated approach with application to FOMC statements* (No. w15367). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w15367/w15367.pdf.

Mahesh, B. (2020). Machine learning algorithms – A review. *International Journal of Science and Research*, *9*(1), 381-386. https://www.ijsr.net/archive/v9i1/ART20203995.pdf.

Oliveira, A. B. and Valls Pereira, P. L. (2018). Uncertainty times for portfolio selection at financial market. Textos para discussão 473, FGV EESP - Escola de Economia de São Paulo, Fundação Getulio Vargas (Brazil). https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/20545/TD%20473_CEQEF%2044.pdf;jsessionid=752FFED35D7A07BADDAEA587BF508616?sequence=1.

Ray, S. (2019). A quick review of machine learning algorithms. In *2019 International conference on machine learning, Big Data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE. https://ieeexplore.ieee.org/document/8862451.

Sagiroglu, S. and Sinanc, D. (2013). Big Data: A review. In *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE.

Sahminan. (2008). Effectiveness of monetary policy communication in Indonesia and Thailand. BIS Working Paper No. 262. https://www.bis.org/publ/work262.pdf.

Schwartz, R, Dodge, J., Smith, N. and Etzioni, O. (2019). Green AI. https://arxiv.org/pdf/1907.10597.

Song, Y., Wang, Y., Ye, X., Wang, D., Yin, Y. and Wang, Y. (2020). Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced credit risk assessment in P2P lending. *Information Sciences*, 525, 182-204.

Strubell, E., Ganesh, A. and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. https://aclanthology.org/P19-1355.pdf.

Sulistiawati, A., Hidayat, M., and Zulen, A.A. (2021). Big Data analytics on RTGS data for tracking corporate activity.

Surjaningsih, N. Werdaningtyas, H. Rahman, F. and Falaqh, R. (2021). Optimalisasi data SLIK sebagai prediktor ketahanan rumah tangga. Laporan Hasil Penelitian Bank Indonesia.

Taylor, J. B. (1999). A historical analysis of monetary policy rules. In *Monetary policy rules* (pp. 319-348). University of Chicago Press.

Teply, P. and Polena, M. (2020). Best classification algorithms in peer-to-peer lending. *North American Journal of Economics and Finance*, 51(C). Doi: 10.1016/j.najef.2019.01.001.

Tissot, B., Doerr, B. S. and Gambacorta, L. (2021). BIS, IFC José María Serena.

Trezecio, M. and Sánchez, J. A. (2019, January). Big Data and machine learning initiatives at the ECB. Bank of Italy and BIS Workshop on Computing Platforms for Big Data and Machine Learning, Rome, 15 January 2019. https://www.bis.org/ifc/events/boibis_jan19/session2A.pdf.

Tumala, M. M. and Omotosho, B. S. (2019). A text mining analysis of central bank monetary policy communication in Nigeria. *Central Bank of Nigeria Journal of Applied Statistics*, *10*(2). https://www.cbn.gov.ng/out/2020/std/73%20-%20107_a419_tumala%20and%20omotosho.pdf.

Wibisono, O., Jabbar, M., Zulen, A. A. and Widjanarti, A. (2021). Machine learning for measuring central bank policy credibility and communication from news.

Wohlfarth, P. (2018). Measuring the impact of monetary policy attention on global asset volatility using search data. *Economics Letters*, *173*, 15-18.

Widjanarti, A., Okfantia, A. and Jabbar, M. (2021). Predicting foreign investors' behaviour and flows projection in Indonesia government bonds market using machine learning. https://www.bis.org/ifc/publ/ifcb57_20.pdf.

Zulen, A. A., Wibisono, O. and Widjanarti, A. (2020). Developing machine learning technique for measuring central bank credibility. *2020 Data for Policy Conference.*

Zulen, A. A., Hidayat, M. and Jabbar, M. (2021). Big Data analytics on payment system data for measuring household consumption in Indonesia.