

IFC-Bank of Italy Workshop on “Data Science in Central Banking: Applications and tools”

14-17 February 2022

Big data analytics on real time gross settlement data for tracking corporate activity¹

Mohammad Khoyrul Hidayat, Amin Endah Sulistiawati and Alvin Andhika Zulen,
Bank Indonesia

¹ This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Big Data Analytics on RTGS Data for Tracking Corporate Activity

Amin Endah Sulistiawati¹, Mohammad Khoyrul Hidayat², Alvin Andhika Zulen³

Abstract

Corporate sector performance is one of the main indicators of financial system stability. Nevertheless, the available information is limited to corporate financial reports which are published a quarterly period with data lag of up to 2 months. The rapid evolution of technology is bringing powerful tools to bear with big data on payment system. This paper explores a new approach for measuring corporate sector activity using Bank Indonesia – Real Time Gross Settlement (BI-RTGS) transaction data by utilizing Big Data Analytics. We identify payment transaction activities (inflow and outflow) via BI-RTGS of more than 100 corporations (with the largest sales) from 9 sectors. We construct a new indicator for tracking corporate sector activity that can be produced on a monthly basis within few days after the reference period. This indicator also has good correlation with sectoral GDP data (for several sectors), which indicates the usefulness of the indicator for tracking economic activity.

Keywords: corporation, payment transaction, RTGS, big data analytics

Introduction

Financial stability is not easy to measure or define as it has complex interconnection of different elements of financial system. The analysis should examine the macroprudential aspects more closely, with more emphasis on the link between the real economy and the financial system. Corporate and household sectors, as well as macroeconomic developments, should be more closely integrated into credit risk assessments of the banking sector (Costeiu Neagu, 2013). By the time, such interactions among financial systems dimensions becomes more complex. The presence of excessive volatility, stress, or crises make financial system characteristics unstable.

Non financial corporations constitute a majority of banking sector's loan borrowers. Corporates' credit quality, loan standards, and lending conditions have significant influence in maintaining healthy economic activity. Fragilities that may

¹ Data Scientist at Digital Data & Big Data Analytics Development Division - Statistics Department, Bank Indonesia

² Data Scientist at Digital Data & Big Data Analytics Development Division - Statistics Department, Bank Indonesia

³ Analyst at Digital Data & Big Data Analytics Development Division - Statistics Department, Bank Indonesia

arise from corporate sector may leap into the banking sector by causing vulnerabilities and systemic risks (Güngör et al., 2016).

Corporate sector performance is one of the main indicators of financial system stability. Based on Bank Indonesia's Financial Stability Review report on 2021, stronger corporate sector performance had a knock-on effect of increasing household income and consumption. However, the available information related to corporate sector performance is limited to financial reports for listed companies. The reports are published in quarterly period with data lag of up to 2 months. Whereas, Bank Indonesia has weekly and monthly Board of Governor meeting. Hence, the need for new more timely indicator for tracking corporate sector activity is becoming more relevant.

On the other hand, the revolution of Big Data in payment system offers richer data provision and techniques to understand and analyse financial system stability. This paper explores a new approach for measuring corporate sector activity using Bank Indonesia - Real Time Gross Settlement (BI-RTGS) transaction data by utilizing Big Data Analytics methodology. We identify payment transaction activities (inflow & outflow) via BI RTGS of more than 100 corporations from 9 sectors. The selected corporations are chosen to represent the largest share in sales for each sector.

Literature Reviews

For a long time, to have access on the state of the economy, policymakers and the private sector have relied on data released by official statistical institutions with usually a lag of several months or years. However, the emergence of technology changes the world and the way it is measured. Based on a recent survey conducted among the members of the Irving Fischer Committee (IFC), the majority of central banks discuss the topic of big data that is used with machine learning applications in a variety of areas, including research, monetary policy and financial stability as well as for supervision and regulation (suptech and regtech applications). Whereas, the quality of data, sampling and representativeness are major challenges for central banks, and so is legal uncertainty around data privacy and confidentiality (Doerr, 2021).

The new era of advanced data analytics is starting to be adopted by the financial sector. Deep learning models nowadays are considered into financial technology stack. (Gensler and Bailey, 2020) used deep learning to explore financial sector and predict financial stability. By exploring 9 characteristics of deep learning (features of hyper-dimensionality, non-linearity, non-determinism, dynamism, and complexity; challenges of explainability, bias, and robustness; and an insatiable hunger for data), it marks a fundamental discontinuity that creates significant opportunities to enhance efficiency, financial inclusion, and risk mitigation. However, extensive adoption of deep learning may also increase uniformity, interconnectedness, and regulatory gaps, which could remain the financial system more fragile.

Another research using big data to analyse the complexity on financial stability is conducted. (Mertzanis, 2018) said that analysis on financial stability and policy should focus on accurate price discovery for complex instruments, realistic financial information generation processes, and system-wide risk materializing within complex financial networks. Thus, complexity analysis can make a useful contribution. In the

past year, the development on standardization on the global legal entity identifier system is made to uniquely identify parties to financial transactions across the globe. This is the first step towards a strong, flexible and adaptable global data infrastructure conducive to financial stability policy.

BI-RTGS Data

Bank Indonesia - Real Time Gross Settlement (BI-RTGS) is an electronic funds transfer system, amongst participants, in rupiah with real-time settlement on an individual transaction basis. BI-RTGS has played an important role in processing payment transactions, including High Value Payment System (HVPS) transactions exceeding Rp100 millions as well as urgent transactions. Data source used in this study is BI-RTGS Transactional Data. We use transaction type code (TTC) 100 and 101 also MT 103 which represents bank's transaction of ordering/beneficiary customer. These type of transaction has approximately 800.000 transactions/month (out of 1,2 millions transaction in total). Since BI-RTGS is in payment infrastructure of Bank Indonesia, we can capture the data on the following day of the settlement time.

Table 1 describes the data structure, which contains Settlement Time, Transaction Type Code, Sender/Receiver Bank Code, Nominal Transaction And Transaction Message (Block 4). While the first 5 fields are structured information, the last one, is in the form of unstructured text. However, Transaction Message field is the fundamental part of this study as it has substance essence information needed.

BI-RTGS Structure Data

Table 1

Settlement Time	Transaction Type Code (TTC)	Sender Bank	Receiver Bank	Nominal	Block 4
Transaction settlement time (MM/DD/YYYY HH:mm)	RTGS transaction code 100 : Transaction o/b of customer 101 : Transaction o/b of customer (w/o account) 111 : Forex Buy/Sell 112 : Interbank Money Market Etc.	Sender's Bank SWIFT Code (BIC) e.g.: CEINAJDJA	Receiver's Bank SWIFT Code (BIC) e.g.: BMRIIDJA	Transaction amount in Indonesian Rupiah	Transaction message, sent by Sender Bank into BI-RTGS system

Transaction Message (Block 4) on data sample, shown in Table 2, is a free text of transaction message filled by sender bank into the system. We highlight on the line text preceded by code 50K for ordering customer information, 59 for beneficiary information and 70K for remittance information. Customer or beneficiary information may consist of account number, corporation name and/or address. Each transaction can have various ways of writing Blok 4 as we do not specify the standard. Hence, to extract this information is another challenge that we will discuss on the next chapter.

BI-RTGS Data Sample

Table 2

Settlement Time	TTC	Sender Bank	Receiver Bank	Nominal	Block 4
5/2/2018 13:35	100	DXXXIDJA	MXXXIDJA	4,000,000,000.00	:20:02RE201804304660#:23B:CRED#:23E:SDVA#:32A:180502IDR4000000000,00# :50K:PAM LYONNAISE JAYA,PT#JL. JEND GATOT SUBROTO KAV 51-52#10270 JAKARTA PUSAT#INDONESIA#:52D:BANK DXXX JAKARTA INDONESIA# :53A:/D/521067000990#DXXXIDJA#:57A:/C/520008000990#MXXXIDJA# :59:/1210005182096#PERSONEL ALIH DAYA, PT# :70:INV. 1804-0167#R/LOCAL# :71A:OUR#:72:/CODTYPTR/100#/CLRC/0670304# :77B:/FEAB/R /PTR/LOCAL
5/2/2018 13:36	100	DXXXIDJA	NXXXIDJA	900,000,000.00	:20:02RE201805020573#:23B:CRED#:23E:SDVA#:32A:180502IDR9000000000,00# :50K:HANJAYA MANDALA SAMPOERNA TBK.PT#RUNKUT INDUSTRI RAYA# NO.18#INDONESIA#:52D:BANK DXXX JAKARTA INDONESIA# :53A:/D/521067000990#DXXXIDJA#:57A:/C/520009000990#NXXXIDJA# :59:/10533549#ASTRA GRAPHIA TBK PT# :70:701625825S,BLACK AND WHITE PRINTED#PAGE#R/LOCAL#: :71A:OUR#:72:/CODTYPTR/100#/CLRC/0670304# :77B:/FEAB/R /PTR/LOCAL

50K : Ordering Customer

59 : Beneficiary Information

70 : Remittance Information

Big Data Analytics Methodology

Big data analytics methodology is used in this study, as we will deal with large amount of payment transaction from BI-RTGS. Not only the volume that is large, but also the velocity is semi-real time. Hereafter, the variety information will be extracted from structure (text and number) and unstructured (free text). Transactional data from BI-RTGS is streamed from payment system infrastructure into Hadoop, an open source framework used for storing and processing big data.

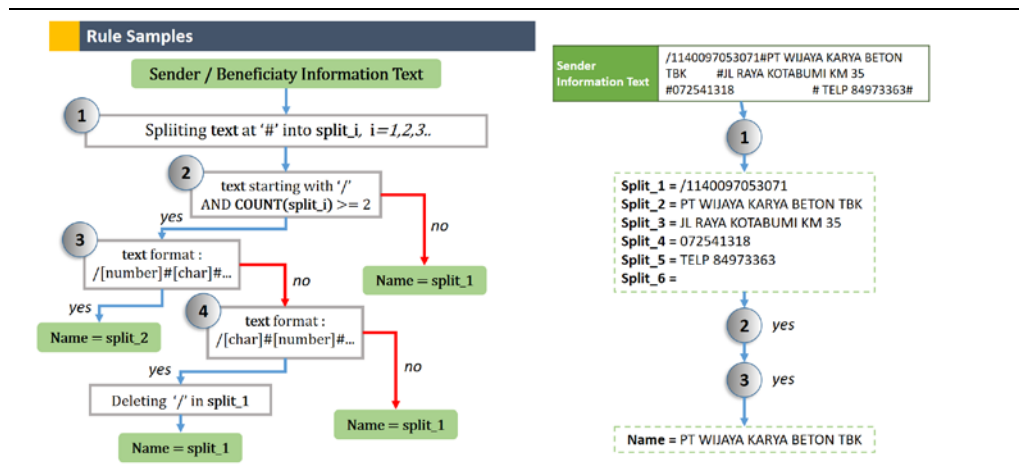
1. Parsing

There are two steps to extract ordering and beneficiary information from Transaction Message (block4) field. The first step is identifying patterns from

message. The second one is defining 61 rules based on identified patten to obtain to corporate name entity. Figure 1 visualize the example of rule based parsing method.

Rule Samples

Figure 1



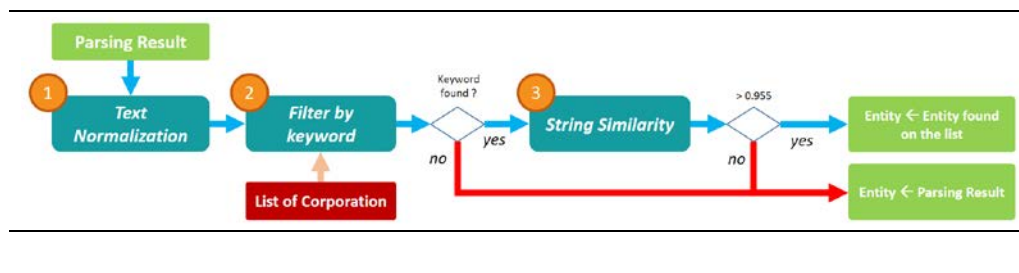
Random sampling of 1.500 transactions data (from 83.300 total transactions on first week period of August 2016) is used for evaluation. Validation on the ruled based parsing result has 96% accuracy. The corporate name entity results is not unique yet because there can be different writing refer to one name. Thus, we need to perform further method to obtain satisfactory result.

2. Entity Resolution

Real-world data, in this case transaction message, is far from perfect due to a plethora of corporate name data entered multiple times in unique sources by various people in their own ways. Entity resolution is a technique used to identify corporate name record in BI-RTGS transactional data that refer to the same corporate and to link the transaction records together. This method, visualized on Figure 1, matches corporate name from transaction message that are nearly identical but maybe not exactly the same.

Identifying similar entities, with different writing

Figure 2



The first step of entity resolution technique is text normalization. We remove non-alphabet characters e.g. dot, comma, space and tab. In order to simplify the writing of corporation name, we replace "and" to symbol "&" and removing stop words such as "PT", "Tbk", "Ltd", "CV", "Pte". The second step is filtering corporations. In this study, we have more than 100 corporation list names. The entity is selected when containing at least 1 (one) word.

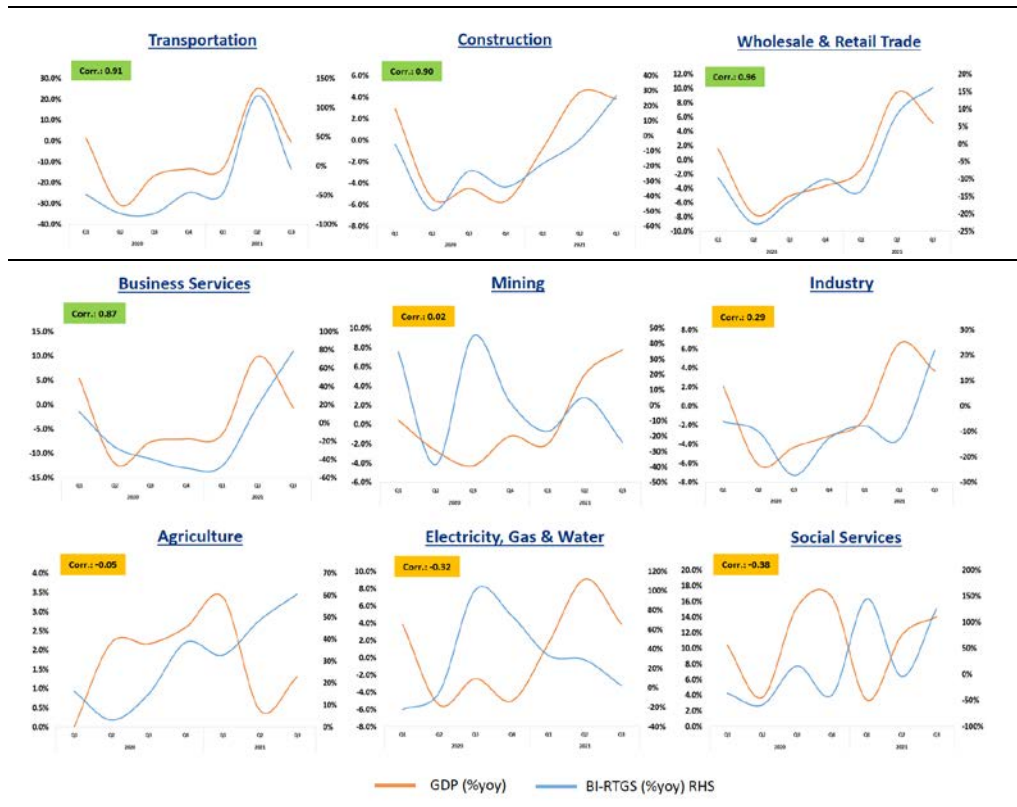
The last one is applying string similarity algorithm. Jaro wringler algorithm is used to calculate entities similarity with threshold more than equal to 0.955. The algorithm is evaluated using random sampling of 1.500 transaction records and results 98.9% accuracy.

Result and Analysis

We construct a new indicator for tracking corporate sector activity that can be produced on a monthly basis within few days after the reference period. We compare corporate sector indicator with sectoral gross domestic product (GDP) data, from Q1 2020 to Q3 2021, published quarterly by The National Statistics Office. The indicators for construction, wholesale and retail trade, transportation and business services sector, have good correlation (≥ 0.87) (Figure 3). Although, mining, industry, agriculture, electricity, gas and water, and social services have lower correlation with GDP.

Corporate sector indicator vs. Indonesian GDP Sector

Figure 3



Conclusion

Big data analytics play an important role to improve the quality of economic analysis and research, as increasingly recognised by policymakers in Bank Indonesia. Applying big data analytics on BI-RTGS transitional data, the new corporate sector indicator is created and (for several sectors) has good correlation with GDP sector. Thus, the indicator can be used for tracking economy activity that will be used as main indicator for financial stability analysis in Bank Indonesia.

References

- S Doerr, L Gambacorta, and J M Serena (2021): "Big data and machine learning in central banking", BIS Working Papers No 930.
- G Gensler and L Bailey (2020): "Deep Learning and Financial Stability", MIT Artificial Intelligence Global Policy Forum.
- C Mertzanis (2018): "Complexity, big data and financial stability", *Quantitative Finance and Economics (QFE)*, 2(3): 637–660.
- G Güngör, M D Özbekler and T P Sümer (2016): "Corporate sector financials from financial stability perspective", Irving Fisher Committee on Central Bank Statistics IFC-ECCBSO-CBRT.
- A Costeiu and F Neagu (2013): "Bridging the Banking Sector With The Real Economy A Financial Stability Perspective", European Central Bank Working Paper Series No 1592.



BANK INDONESIA
BANK SENTRAL REPUBLIK INDONESIA



Big Data Analytics on Real Time Gross Settlement Data for Tracking Corporate Activity

Amin Endah Sulistiawati, Mohammad Khoyrul Hidayat, Alvin Andhika Zulen
Statistics Department – Bank Indonesia
Email: amin_endah@bi.go.id, moh_khoyrul@bi.go.id, alvin_az@bi.go.id

February 2022

*The views expressed here are those of the authors and
do not necessarily reflect the views of Bank Indonesia*



OUTLINE

1 Background

2 Proposed Idea

3 Data

4 Methodology

5 Result & Analysis





Background

Non-financial corporations constitute a majority of banking sector's loan borrowers. Corporates' credit quality, loan standards, and lending conditions have significant influence in maintaining healthy economic activity. **Fragilities that may arise from corporate sector may leap into the banking sector by causing vulnerabilities and systemic risks** (Güngör et al., 2016).

Financial stability analyses should examine the macroprudential aspects more closely, with more emphasis on the link between the real economy and the financial system. Corporate and household sectors, as well as macroeconomic developments, should be more closely integrated into credit risk assessments of the banking sector (Costeiu & Neagu, 2013).

Corporate sector performance is one of the main indicators of financial system stability. Based on Bank Indonesia's Financial Stability Review report on 2021, stronger corporate sector performance had a knock-on effect of increasing household income and consumption.

The available information related to corporate sector performance is limited to financial reports for listed companies. The reports are published in quarterly period with data lag of up to 2 months. Whereas, Bank Indonesia has weekly and monthly Board of Governor meeting. Hence, **the need for new & more timely indicator for tracking corporate sector activity is becoming more relevant.**

Proposed Idea

This paper explores a new approach for measuring corporate sector activity using Bank Indonesia – Real Time Gross Settlement (BI-RTGS) transaction data by utilizing Big Data Analytics methodology.

We identify payment transaction activities (inflow & outflow) via BI-RTGS of more than 100 corporations from 9 sectors. The selected corporations are chosen to represent the largest share in sales for each sector.



Agriculture



Mining



Industry



Electricity, gas, & water



Construction



Trade



Transportation



Business services



Social services



Data Source : BI-RTGS Transactional Data
Transaction Type : Banks' transactions o/b of customer (TTC 100 & 101) ~ MT103
Total Transactions : $\cong$ 800.000 transactions/months (out of 1,2 mio transactions)

BI-RTGS



DATA STRUCTURE

Settlement Time	Transaction Type Code (TTC)	Sender Bank	Receiver Bank	Nominal	Block4
Transaction settlement time (MM/DD/YYYY HH:mm)	RTGS transaction code 100 : Transaction o/b of customer 101 : Transaction o/b of customer (w/o account) 111 : Forex Buy/Sell 112 : Interbank Money Msrket etc.	Sender's Bank SWIFT Code (BIC) e.g.: CENAI DJA	Receiver's Bank SWIFT Code (BIC) e.g.: BMRI DJA	Transaction amount in Rupiah	Transaction message, sent by Sender Bank into BI-RTGS system

Settlement Time	TTC	Sender Bank	Receiver Bank	Nominal	Block4
5/2/2018 13:35	100	DXXXIDJA	MXXXIDJA	4,000,000,000.00	:20:02RE201804304660#:23B:CRED#:23E:SDVA#:32A:180502IDR4000000000,00# :50K:PAM LYONNAISE JAYA,PT#JL. JEND GATOT SUBROTO KAV 51-52#10270 JAKARTA PUSAT#INDONESIA# :52D:BANK DXXX JAKARTA INDONESIA# :53A:/D/521067000990#DXXXIDJA#:57A:/C/520008000990#MXXXIDJA# :59:/1210005182096#PERSONEL ALIH DAYA, PT# :70:INV. 1804-0167#R/LOCAL# :71A:OUR#:72:/CODTYPTR/100#/CLRC/0670304#:77B:/FEAB/R /PTR/LOCAL

50K: Ordering Customer

59: Beneficiary Information

70: Remittance Information

Methodology

BI-RTGS



Hadoop



1. Parsing

Extracting ordering and beneficiary entity name from transaction message (block4)



2. Entity Resolution

Identifying similar entities, with different writing



3. Sectoral Mapping

Grouping corporations by sector



Data Extraction



4. Validation & Analysis

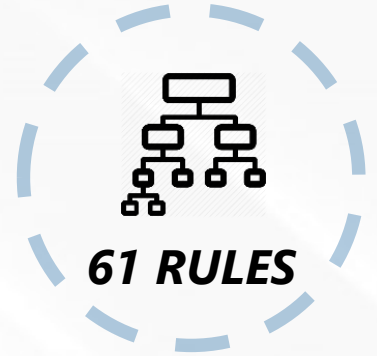
- We aggregate the data for incoming & outgoing transactions by sector.
- We calculate the correlation between the new indicator with sectoral GDP data.



Extracting sender & beneficiary information from “Transaction Message (block4)” field

Method

- ✓ Identifying writing patterns for sender & beneficiary information .
- ✓ Define some rules in order to extract entities name in sender/beneficiary column based on writing patterns and sender bank code.

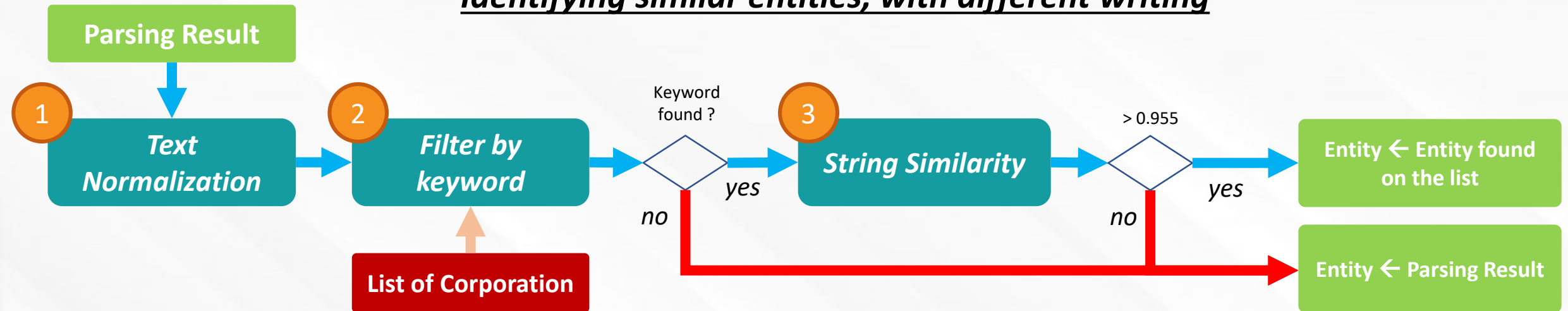


Pattern Samples

RTGS Data (Sender / Beneficiary Information)	Pattern
/NO-ACC#PT SARANA INDAH PERKASA#	/NO-ACC#name#
/NO-ACC#PT LONTAR PAPYRUS PULP DAN PAPER IN#DUSTRY#	/NO-ACC#name#name#
/NO-ACC#PT PELABUHAN INDONESIA IV (PERSERO)# MAKASSAR#	/NO-ACC#name#address#
/NO-ACC#317104000100402170/ 071609261406260#001#	/NO-ACC#acc. number#acc. number#
/703096771400#SATYA MITRA SEJAHTERA#JALAN DIPONEGORO NO	/acc.number#name#address
/823000009900#ANEKA SAWIT LESTARI#THE PLAZA OFFICE TOWER 35TH FLOOR#JL MH THAMRIN KAV 28-30#10350#	/acc.numbe#name#address#address#address#
SWASTISIDDHI AMAGRA PT#	name#
/PT GEMA GRAHA SARANA TBK#09102878785#	/name#acc. number#

Methodology: Entity Resolution

Identifying similar entities, with different writing



1. Text Normalization

- Removing *non-alphabet* character, e.g. “.” (dot) , “,” (comma) , “ ” (space), “\t” (tab)
- Replacing “and” & “dan” with “&”
- Removing stop words, e.g.: “PT”, “Tbk”, “Ltd”, “CV”, “Pte”

2. Filter by keyword

Filtering corporations from the list, containing at least 1 (one) word from text normalization result.

3. String Similarity

Using Jaro-Winkler algorithm to calculate similarity between entities.

Criteria : String similarity ratio ≥ 0.955

Result & Analysis

We construct a **new indicator for tracking corporate sector activity** that can be produced **on a monthly basis within few days after the reference period**. This indicator also has good correlation with sectoral GDP data (for several sectors), which indicates the usefulness of the indicator for tracking economic activity.

Construction



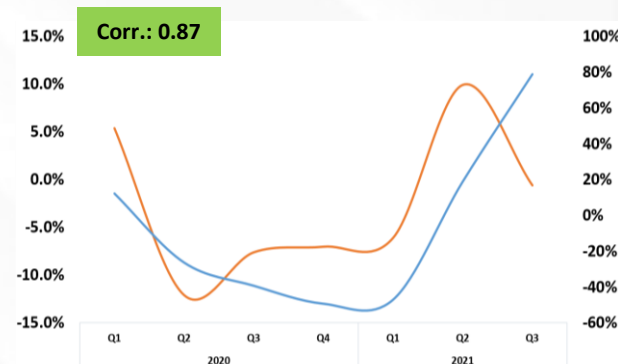
Wholesale & Retail Trade



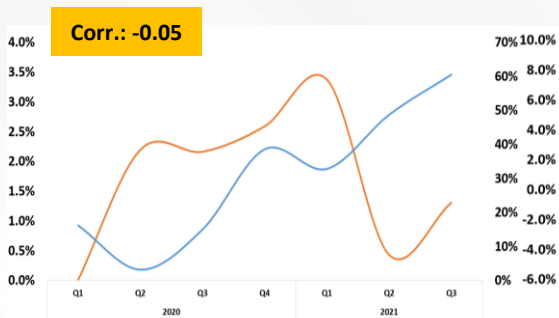
Transportation



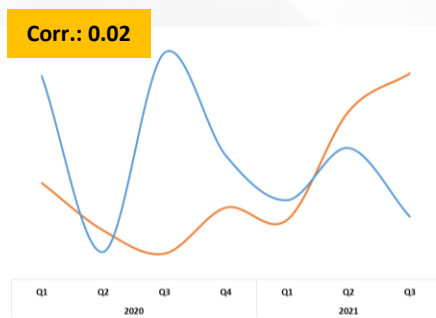
Business Services



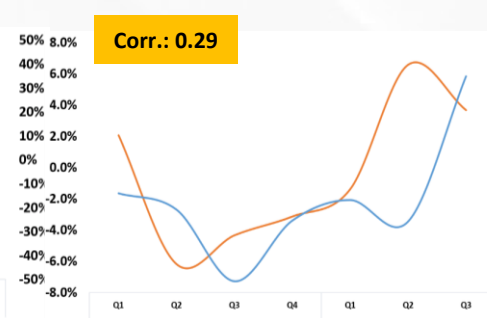
Agriculture



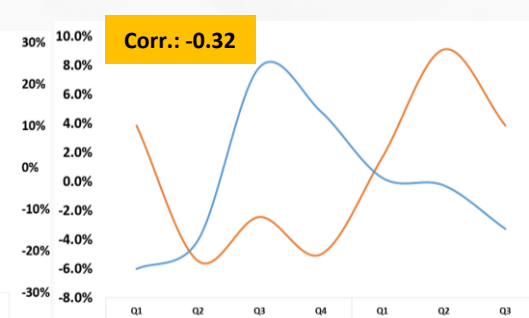
Mining



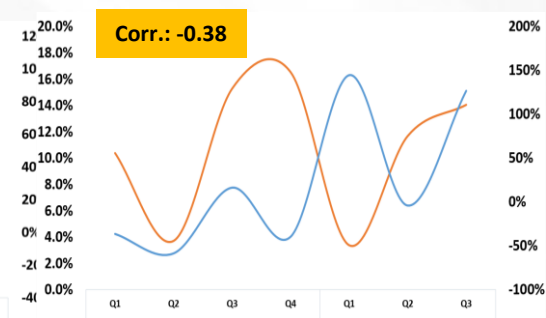
Industry



Electricity, Gas & Water



Social Services



— Sectoral Real GDP (%yoy) — Total Inflow + Outflow BI-RTGS (%yoy) RHS



BANK INDONESIA
BANK SENTRAL REPUBLIK INDONESIA

THANK YOU TERIMA KASIH

Amin Endah Sulistiawati, Mohammad Khoirul Hidayat, Alvin Andhika Zulen
Statistics Department – Bank Indonesia
Email: amin_endah@bi.go.id, moh_khoirul@bi.go.id, alvin_az@bi.go.id