11th Biennial IFC Conference on "Post-pandemic landscape for central bank statistics"

BIS Basel, 25-26 August 2022

Big data analytics on payment system data for measuring household consumption in Indonesia[1]

Renardi Ardiya Bimantoro, Mohammad Khoyrul Hidayat,
Muhammad Abdul Jabbar and Alvin Andhika Zulen,
Bank Indonesia

[1] This presentation was prepared for the conference. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the event.

Big Data Analytics on Payment System Data for Measuring Household Consumption in Indonesia

Alvin Andhika Zulen [1], Mohammad Khoyrul Hidayat [2], Muhammad Abdul Jabbar[3], Renardi Ardiya Bimantoro[4]

Abstract

Consumer spending is one of the main indicators to measure state of the economy in Indonesia. However, those data are published on a quarterly basis with a publication lag of one month. This study examines the use of retail payment system data as a proxy for household consumption indicators in Indonesia. By utilizing Big Data Analytics methodology, we are able to construct an indicator, which is available within a few days after the end of the reference period. This indicator can be used as initial proxy for household consumption, which is indicated by a good correlation with the official data.

Keywords: GDP; household consumption; payment system; big data

JEL classification: B22, C55, E21

[1] Statistics Department – Bank Indonesia; e-mail: alvin_az@bi.go.id

[2] Statistics Department – Bank Indonesia; e-mail: moh_khoyrul@bi.go.id

[3] Statistics Department – Bank Indonesia; e-mail: muhammad_abdul@bi.go.id

[4] Statistics Department – Bank Indonesia; e-mail: renardi_ardiya@bi.go.id

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

Contents

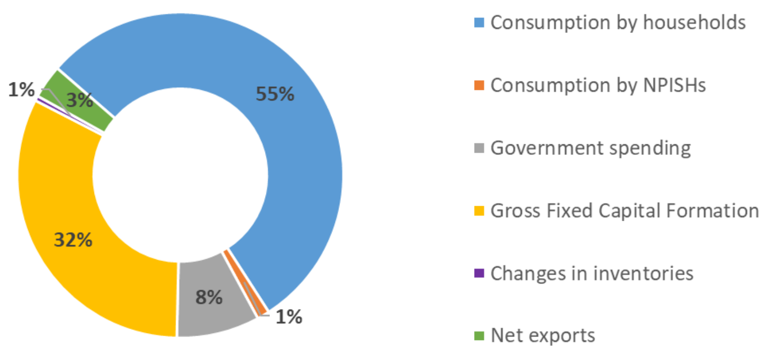
1. Background

The growth of an economy can be measured by Gross Domestic Product (GDP) data. GDP can be calculated through three approaches, i.e. production, income, and expenditure approach. The expenditure approach is a crucial aspect for Bank Indonesia in carrying out its mandate as the monetary, macroprudential, and payment system authority.

The expenditure approach can be analysed through several indicators. One of the indicators is the household consumption expenditure indicator, which represents the spending on goods and services by resident households for final consumption purposes. Household consumption is the most significant contributor (±55%, Figure 1) to Indonesia's GDP based on the expenditure in 2020.

Contribution of Indonesia's GDP Components by Expenditure in 2020

Figure 1



Source: BPS

Bank Indonesia, in synergy with the government, constantly strives to formulate the appropriate policies in its 3 (three) main objectives of Bank Indonesia, which are monetary, financial system stability, and payment system. For the central bank, it is significant to know the current economic condition (state of the economy) which will be used as the basis for predicting the economy's growth in the future. Considering Bank Indonesia's policies are aimed at influencing the expenditure sides, Bank Indonesia needs to observe the movement of household consumption as the most significant expenditure component in Indonesia's GDP as early as possible. However, data related to household consumption indicators in GDP are published and available on a quarterly basis with a publication lag of one month.

Nowcasting is a method used to predict the direction of the economic movement. Through nowcasting, policymakers can assess the direction of the economic movement by using representative high-frequency data to capture the dynamics of the reference indicators (i.e. GDP). Many researches related to nowcasting have been published, and several nowcasting models have been widely used to estimate various indicators (Tarsidin, Idham, & Rakhman, 2016). In addition, this method may help policymakers to formulate policy responses while waiting for the official release of macroeconomic indicators.

The Covid-19 pandemic since the beginning of 2020 has directly impacted the world economy, and Indonesia is no exception. Indonesia has fallen into its first recession in 22 years as the Covid-19 pandemic continues to take it toll. In response to this situation, policymakers need to project macroeconomic indicators to formulate appropriate policies. During this pandemic, the analysis of household consumption indicators as one of the macroeconomic indicators is very fundamental, especially in helping the central bank and government to see the growth and predict household consumption behavior.

On the other hand, technological advancement and the widespread use of cashless payment systems in the digital era have opened up the opportunities to explore large dataset of payments data for monitoring economic activity. For example, current technological advancement allow us to utilize Big Data Analytics for processing large dataset and estimating economic indicators in advance (Buono et al., 2018). This study aims to examine the use of retail payment system data, particularly from the Bank Indonesia National Clearing System (SKNBI), which has a high availability frequency, as a proxy for household consumption indicators in Indonesia.

2. Literature Review

2.1 Utilization of Big Data for Macroeconomic Indicators

Along with the technological advancement, various parties have taken the advantage of Big Data Analytics more broadly. To be more specific, near real-time and faster data processing is urgently needed, especially for supportingp policy formulation during the current Covid-19 pandemic. By collecting and processing data on a large scale and high frequency, central bank can determine the current state of the economic in advance as a basis for policy formulation. In addition, literature studies related to the use of Big Data Analytics in the economic field have been developed with various methodologies.

Kapetanios & Papailias (2018) discusses the potential use of Big Data Analytics in nowcasting GDP and other macroeconomic indicators in the UK. In this study, the authors describe various initiatives related to Big Data Analytics in nowcasting macroeconomic indicators. The research also describes the benefit of using Big Data Analytics, which makes it possible to process monthly, weekly, daily, or higher frequency data on a large scale.

In another study, Buono et al. (2018) discusses GDP projection by utilizing various type of data: (i) macroeconomic data with monthly frequency, i.e., core consumer prices, consumer price index, house prices, job vacancies index; (ii) financial data with weekly frequency, i.e., Interest rates, equity indexes, and (iii) uncertainty indicators based on keyword searches in Google. In this study, the author concludes that the results of the uncertainty indicator from Big Data contribute in reducing the RMSFE (root mean squared forecast error) between the nowcasting results and the actual value.

2.2 Utilization of Payment System Data

Several studies have proven that high-frequency data, i.e., data from the payment system, can be used to estimate macroeconomic indicators. By utilizing high-frequency data, which is available faster, we can produce an earlier estimate of current economic conditions. Through this approach, the policy-making authorities benefit by being able to obtain prompt indicators for supporting policy assessment and formulation.

For example, Galbraith & Tkacz (2015) conducted an assessment in using large-scale datasets from the payment system, i.e., debit card, credit card, and cheque transactions, as a proxy for GDP growth in Canada. This study found that by using payment system data as one of the input variables can reduce the nowcasting error by 65%, compared to only using macroeconomic indicators as the input variables.

Recent research was also conducted by Dunn et al. (2020) to measure the impact of the Covid-19 pandemic on consumer spending by utilizing payment transaction data. This study shows a high correlation between official survey data and payment transaction data, especially for retail, accommodation, and restaurant sectors. In terms of data availability, the payment transaction data can be available daily with a lag of 3 (three) days, much higher in frequency when compared to data from monthly surveys that have a publication time lag of 1 (one) month. This study concludes that payment transaction data can be used as alternative data and initial proxy for consumption indicators.

Thus, this study is expected to answer the following research questions:

1. Can payment system data be used as a data source to measure household consumption indicators in Indonesia?
2. Can the resulting household consumption indicators complement the existing indicators?

3. Methodology

3.1 Data

The data source used in this study was obtained from the National Clearing System of Bank Indonesia (SKNBI). This National Clearing System of Bank Indonesia (SKNBI) is a Retail Value Payment System (RVPS) infrastructure operated by Bank Indonesia to process electronic financial data for fund transfer services, debit clearing services, regular payment services, and regular billing services (Regulation of Member of Board of Governors Number 21/12/PADG/2019). Since September 2019, SKNBI has been able to process payment transactions with less than Rp. 1,000,000,000.00 (one billion rupiah) in amount.

The scope of data used in this study is SKNBI fund transfer transaction data from July 2015 to December 2021. Fund transfer service is a service within SKNBI that facilitates the transfer of funds between participant banks, with an average number of transactions reaching ±13 million transactions per month. In the fund transfer service, there are several types of transaction:

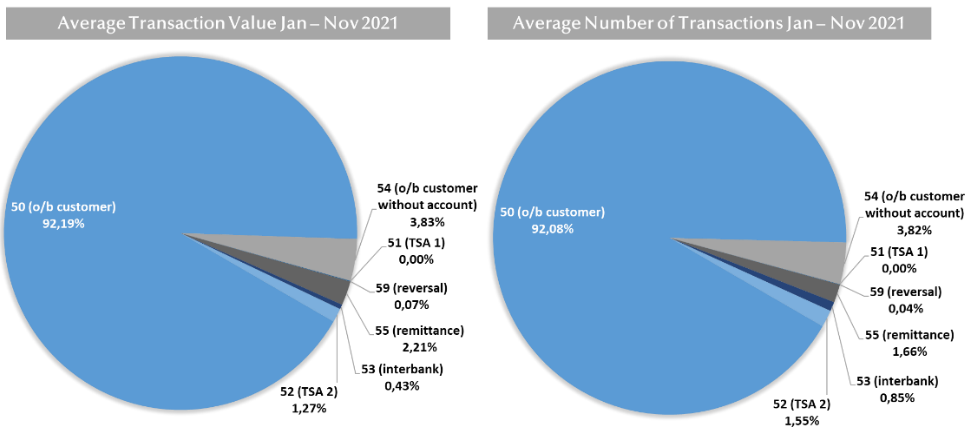
1. 50: Transfer of funds between participants on behalf of the customer;

2. 51: Transfer of funds between participants related to Government's Treasury Single Account (TSA);
3. 52: Transfer of funds between participants and Bank Indonesia's Treasury Single Account;
4. 53: Transfer of funds between participants that is not for the customers' needs;
5. 54: Transfer of funds between participants on behalf of the customer without accounts;
6. 55: Transfer of funds between participants on behalf of the customer related to money remittance; and
7. 59: Refund of fund transfers and payments (reversal).

The share for each transaction type is shown in Figure 2, which shows that the largest share of transaction (±92%) is fund transfer transactions between participants on behalf of the customer (code 50).

Nominal and Frequency of SKNBI Transactions Based on Transaction Type

Figure 2



Source: SKNBI (processed)

The data structure obtained from the fund transfer service transactions is as shown in Table 1. Although the data structure of the SKNBI fund transfer is quite comprehensive, there are still issues related to the data validity. For example, we found that the customer's type code does not always match the customer's name category, both in the sender and recipient fields.

SKNBI Fund Transfer Data Structure

Table 1

No.	Data Field
1	DKE ID
2	BATCH ID
3	TRANSACTION DATE
4	SENDER BANK CODE
5	SENDER LOCATION
6	BENEFICIARY BANK CODE
7	BENEFICIARY LOCATION
8	AMOUNT
9	TRANSACTION TYPE CODE
10	SENDER CUSTOMER'S NAME
11	SENDER CUSTOMER'S ACCOUNT NUMBER

No.	Data Field
12	SENDER CUSTOMER'S ADDRESS
13	SENDER CUSTOMER'S ID NUMBER
14	SENDER CUSTOMER'S TYPE
15	BENEFICIARY CUSTOMER'S NAME
16	BENEFICIARY CUSTOMER'S ACCOUNT NUMBER
17	BENEFICIARY CUSTOMER'S ADDRESS
18	BENEFICIARY CUSTOMER'S ID NUMBER
19	BENEFICIARY CUSTOMER'S TYPE
20	DESCRIPTION

3.2 Workflow

In constructing household consumption indicators from SKNBI fund transfer data, we develop a text mining model with a rule-based approach for processing unstructured information, backed up by parallel computing technology in Apache Spark – Hadoop. We use Python as the programming language. In general, the workflow in this study consists of data preprocessing, data extraction, and data validation.

3.2.1 Data Preprocessing

The data preprocessing stage is carried out to prepare the raw SKNBI fund transfer transaction data so that they can be further processed at the next stage. The process is as follows:

1. Filter transactions that are not on behalf of customers.

The data that will be further processed is only data with transaction type code 50 (transfer of funds between participants on behalf of the customer) and 54 (transfers of funds on behalf of the customer without accounts).

2. Classification of customers' categories.

For each transaction, we classify sender and beneficiary customers into business entities, governments, and others. This process is critical since customers' type code has validity issues. Classification is conducted using a rule-based approach with rules as shown in Table 3.

3. Filter transactions that do not have a description.

Rules for Classification of Customers' Categories — Table 3

CUSTOMER CLASSIFICATION	SAMPLE KEYWORDS
Business Entities	*'PT.', 'CV.', UD.', 'PD.', 'BANK', 'TBK', 'PERUSAHAAN', 'INTERNATIONAL', 'INTERNASIONAL', 'FINANCE', 'LIFE', 'INSURANCE'.*
Government	*'OTORITAS', 'BADAN', 'BPJS' 'DINAS', 'KAB.', 'KABUPATEN', 'KECAMATAN', 'KELURAHAN', 'KPU', 'PEMERINTAH', 'PEMKAB', 'PEMKOT', 'PROVINSI', 'PROV.', 'PEMPROV'.*
Others (individual)	Does not contain business entities and governments keywords.

3.2.2 Data Extraction

There is limited information on the description of fund transfer transactions in SKNBI, in which there is no standard format/reference code for the information written in the description field (free text). We develop a text mining model to analyze the transaction data to handle this issue. The resulting data from the previous stage (section 3.2.1) are used as input for this stage with the following process:

1. **Classification of the purposes of SKNBI fund transfer transactions.**

   Classification is done using a rule-based approach based on predefined keywords, as shown in Table 4.

2. **Data aggregation.**

   After classifying the purpose of the transaction, data can be aggregated as indicators for each transaction purposes, e.g. household consumption, household income, and business. As for transactions with transfer purposes other than those three categories are classified as "Others".

Rules for Classification of Transaction Purposes — Table 4

SENDER	BENEFICIARY	KEYWORD IN TRANSACTION DESCRIPTION	CLASSIFICATION
Other than Business Entities and Government	-	'pembayaran', 'konsumsi', 'belanja', 'pelunasan', 'kos', 'cicilan', 'dp', 'payment', 'dana', 'pelunasan', 'setoran', 'angsuran', 'jasa', 'invoice', 'listrik', 'service', 'sewa rumah', 'kasbon', dsb.	Household Consumption
-	**Other than Business Entities and Government**	'gaji', 'honor', 'upah', 'payroll', 'salary', 'remunerasi', 'insentif', 'lembur', 'overtime', ''kompensasi', 'bagi hasil', 'bonus', 'komisi', 'tukin', 'uang makan',dsb.	Household Income
Business Entities	**Business Entities**		Business

3.2.3 Data Validation

The resulting indicators from the previous stage (section 3.2.2) are then validated with the GDP data - Household Consumption (current prices) as the reference indicator. The monthly indicators from the SKNBI are converted into quarterly data by accumulating the nominal amount of transactions in each quarter. This step is required to obtain indicators with the same frequency as household consumption data in GDP. After that, validation is conducted by calculating the correlation value between those two data.

4. Result and Analysis

4.1 Evaluation of Classification Model

After we develop the classification model with the rule-based approach in the previous section, we need to evaluate our model to find out how accurate the model is in classifying the transaction purposes, particularly for consumption. In this study, we use F1-score[5] as an evaluation metric. The evaluation was carried out on ±4.000 transactions (random sampling) during the period of 2018 to 2021.

The evaluation results in Table 5 show us a good value of the overall F1-score (84,5%). However, the F1-score for predicting consumption transaction is still relatively low compared to the results for other categories. If we analyze using the confusion matrix in Table 6, there are still plenty of false positive cases, i.e. transactions predicted to be "consumption" but should be included in other categories. We suspect that the prediction error could be caused by the accuracy of the customer categories classification, which still need to be improved.

Evaluation of Transaction Purposes Classification Model — Table 5

TRANSACTION PURPOSES	*RECALL*	*PRECISION*	F1-*SCORE*	AVERAGE OF F1-*SCORE*	ACCURACY
Consumption	80,2%	62,6%	70,3%	**84,5%**	**88,3%**
Income	85,4%	98,8%	91,6%		
Business	75,6%	98,2%	85,4%		
Others	97,4%	84,5%	90,5%		

Confusion Matrix of Transaction Purposes Classification Model — Table 6

		PREDICTION			
	Transaction Purposes	**Consumption**	**Income**	**Business**	**Others**
ACTUAL	**Consumption**	**154**	**0**	**0**	**38**
	Income	**17**	**569**	**4**	**76**
	Business	**42**	**0**	**1.078**	**240**
	Others	**33**	**7**	**12**	**1.935**

4.2 Result Validation

As previously explained in section 3.2.3, the household consumption indicator from the SKNBI (growth, y.o.y) is validated with the GDP-household consumption indicator at current prices (growth, y.o.y). The correlation of the two indicators can be seen in Table 7 and the graph visualization in Figure 4.

The validation results show a high correlation between the two indicators since 1st quarter of 2019, including during the Covid-19 pandemic. These results indicate that consumption indicators from SKNBI transaction data can be used as a proxy for household consumption. Moreover, it can be available earlier, i.e. 2 (two) days lag after the end of the period, both weekly and monthly. The availability of this indicator

[5] *F1* : 2 x *precision* x *recall* / (*precision* + *recall*)
Precision : (*true positive*) / (*true positive* + *false positive*)
Recall : (*true positive*) / (*true positive* + *false negative*)

is much faster than the publication of GDP data which has a time lag of more than 1 (one) month.

Correlation Table

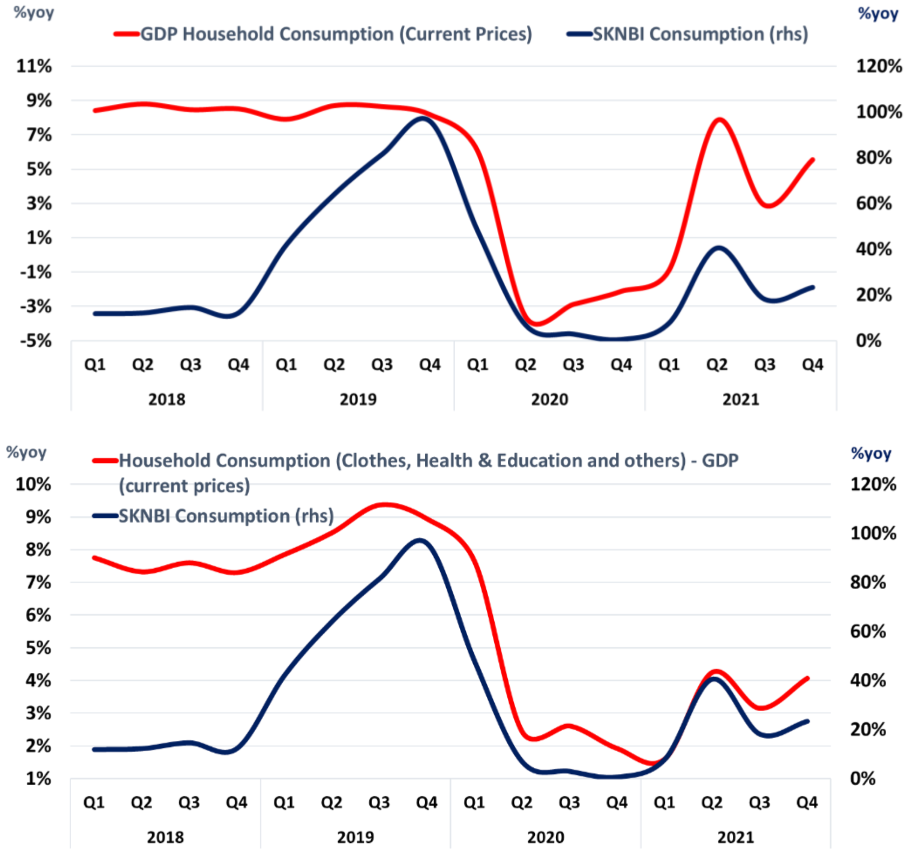
Table 7

INDICATORS	CORRELATION OF SKNBI CONSUMPTION WITH GDP HOUSEHOLD CONSUMPTION (GROWTH, YOY)		
	Q1-2018 to Q4-2021	Q1-2019 to Q4-2021	Q1-2020 to Q4-2021
Total of Household Consumption	55,2%	85,2%	91,2%
Clothes, Health & Education, Restaurant & Hotel, and Others	61,3%	85,2%	90,4%
Clothes, Health & Education, and Others	67,7%	93,3%	90,3%
Health & Education	65,9%	83,4%	60%
Restaurant & Hotel	50,8%	69,2%	80,2%

Source: SKNBI, BPS (processed)

SKNBI Consumption Growth and Household Consumption Growth – GDP (percent, y.o.y)

Figure 4



Source: SKNBI, BPS (processed)

5. Conclusion and Future Work

5.1 Conclusion

In this study, we have proposed a new approach in utilizing payment system transaction data as a proxy for household consumption indicators. Using text mining methodology with rule-based model, we can classify customer categories and transaction purposes from the SKNBI fund transfer data. Based on the evaluation of the model, the average F1-score of the model is 84,5%.

Using this methodology, we can obtain a proxy for household consumption indicators from high-frequency payment system data more quickly, compared to the official publication of GDP data. The validation results show a high correlation between these two indicators, which indicates that the household consumption indicator from the SKNBI fund transfer data can be used as a proxy for household consumption indicators.

5.2 Future Work

There are several improvements in the methodology that can be applied for future works.

1. Improving the methodology for classifying customers' categories and transaction purposes, including the use of machine learning algorithms.
2. Utilizing Bank Indonesia's Fast Payment (BI-FAST) transaction data (implemented since the end of 2021) as additional data source in constructing consumption indicators from retail payment system.
3. Using consumption indicators from the payment system, e.g. fund transfers from SKNBI and payment transactions via cards, with other macroeconomic variables, to construct the nowcasting model of household consumption.

References

Buono, D., Kapetanios, G., Macellino, M., Mazzi, G., & Papailias, F. (2018). *Big Data Econometrics: Now Casting and Early Estimates*. Universita Bocconi.

Dunn, A., Hood, K., & Driessen, A. (2020). *Measuring the Effects of the COVID-19 Pandemic on Consumer Spending Using Card Transaction Data*. Bureau of Economic Analysis.

Galbraith, J. W., & Tkacz, G. (2015). *Nowcasting GDP with Electronic Payments Data*. European Central Bank.

Bank Indonesia. *Peraturan Anggota Dewan Gubernur Bank Indonesia Nomor 21/12/PADG/2019 tentang Penyelenggaraan Transfer Dana dan Kliring Berjadwal oleh Bank Indonesia*.

Kapetanios, G., & Papailias, F. (2018). *Big Data & Macroeconomic Nowcasting: Methodological Review*. The Economic Statistics Centre of Excellence.

Tarsidin, Idham, & Rakhman, R. N. (2016). *Nowcasting Konsumsi Rumah Tangga dan Investasi*. Bank Indonesia.

BANK INDONESIA
BANK SENTRAL REPUBLIK INDONESIA

Big Data Analytics on Payment System Data for Measuring Household Consumption

Alvin Andhika Zulen, Mohammad Khoyrul Hidayat, Muhammad Abdul Jabbar, Renardi Ardiya Bimantoro
Statistics Department – Bank Indonesia
Email: alvin_az@bi.go.id, moh_khoyrul@bi.go.id, muhammad_abdul@bi.go.id, renardi_ardiya@bi.go.id

August 2022

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

OUTLINE

1. **Background**
2. **Data Source**
3. **Methodology**
4. **Result & Analysis**
5. **Conclusion**

BACKGROUND

BACKGROUND

- **Household consumption is one of the main indicators to measure state of the economy in Indonesia** (largest contributor, 55%, in Indonesia's GDP). However, GDP data (incl. household final consumption expenditure) are published and available on a quarterly basis with a publication lag of one month.
- Bank Indonesia provides retail value payment system, SKNBI (The National Clearing System), that can **generate data related to fund transfers, including household transactions**.
- Advancements of technology and widespread use of payment systems have opened the **opportunity to explore large dataset of payment data for monitoring economic activity**.

OBJECTIVE

Developing a high frequency measure of household consumption in Indonesia **from retail value payment system data** (SKNBI), by utilizing Big Data Analytics methodology, particularly text mining.

DATA SOURCE

DATA SOURCE

Fund Transfer of SKNBI:

Credit transfer transaction between participants (banks) on behalf of the customers.

- ✓ **Total Transactions** ≅ 13 mio trx/month
- ✓ **Nominal Transactions** ≤ Rp. 1 Billion/trx
- ✓ **Availability Period** : July 2015 s.d. December 2021

DATA STRUCTURES

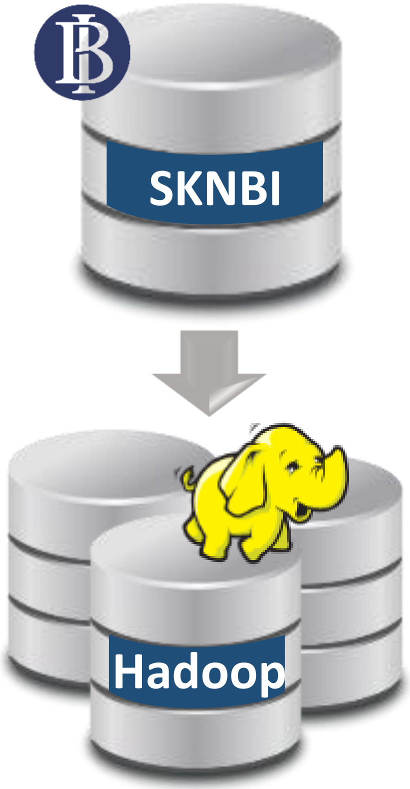
	COLUMN NAME
1	DKE ID
2	BATCH ID
3	TRANSACTION DATE
4	ORIGINATING BANK CODE
5	SENDER LOCATION
6	BENIFICIARY BANK CODE
7	RECEIVER LOCATION
8	AMOUNT
9	TRANSACTION TYPE CODE

	COLUMN NAME
10	SENDER CUSTOMER’S NAME
11	SENDER CUSTOMER’S ACC NUMBER
12	SENDER CUSTOMER'S ADDRESS
13	SENDER CUSTOMER'S ID NUMBER
14	SENDER CUSTOMER’S TYPE CODE
15	BENEFICIARY CUSTOMER’S NAME
16	BENEFICIARY CUSTOMER’S ACC NUMBER
17	BENEFICIARY CUSTOMER'S ADDRESS
18	BENEFICIARY CUSTOMER'S ID NUMBER
19	BENEFICIARY CUSTOMER’S TYPE CODE
20	DESCRIPTION

METHODOLOGY

OVERALL WORKFLOW

Data Acquisition



Data Preprocessing

1

- Filtering type of transaction.
- Classification of customer categories into household and business entities.
- Removing any transactions without description.

Data Extraction

2

- Classification of transactions into consumption, income, and other trasactions (using rule-based*).
- Data aggregation and constructing consumption and income indicator.

Data Validation

3

Validation with official indicators
(i.e. GDP – household consumption)

*) e.g. : Consumption is SKNBI Fund Transfer with transaction detail containing keywords related to household consumption, e.g .: 'buy', 'shop', 'pay', 'paid off', 'installments', etc.

METHODOLOGY – DATA PREPROCESSING

We use rule-based (keyword) approach for classifying customer categories.

Keywords	Category
'PT.', 'CV.', UD.', 'PD.', 'BANK', 'TBK', 'PERUSAHAAN', 'SDN BHD', 'INTERNATIONAL', 'INTERNASIONAL', 'FINANCE', 'LIFE', 'UNITED', 'RESTAURANT', 'ASURANSI', 'INSURANCE', 'AUTOMOTIVE', 'MANUFACTURING', 'MOTOR', 'LOGISTIC', 'LOGISTIK', 'PRODUCT', 'DEVELOPMENT', 'INDUSTRY', 'INDUSTRIES', 'PHARMA', 'CORP', 'PROD', 'DAIRY', 'LTD', 'GRAND', 'HOTEL', 'GREEN', 'CONSULTING', 'GROUP', 'GLOBAL', 'INTER', 'AGRO', 'RESORTS', 'TRANS', 'JASA', 'AJB', 'SECURITY', 'TEXTILE', 'INDO', 'SUKSES', 'RUMKIT', 'SUMBER', 'PERTAMINA', 'PLN', 'JASAMARGA', 'ELECTRONIC', 'INTL', 'POSCO ', 'PTPN ', '.CO', 'KANTOR PUSAT', 'PERUM ', 'PROPERTY ', 'RESEARCH', 'LIMITED', 'MEDICAL ', 'ASTRA ', 'HUSADA', 'TRADE ', 'AGRICULTURE ', 'FOOD', 'OPERATION', 'MAINTENANCE', 'SERVICES', 'STEEL', 'SUPPLIES'	**Business Entities**
'OTORITAS', 'RSUD', 'BADAN', 'BPJS', 'PUSKESMAS', 'DINAS', 'KAB.', 'KABUPATEN', 'KEC.', 'KECAMATAN', 'KEL.', 'KELURAHAN', 'KPU', 'PEMERINTAH', 'PEMKAB', 'PEMKOT', 'PROVINSI', 'PROV.', 'PEMPROV', 'RKUD', 'EMBASSY', 'DITJEN', 'DIKBUD', 'CAMAT', 'DAERAH', 'PDAM', 'KERETA', 'KAS UMUM', 'RPKBUN', 'RPK-BUN', 'SPAN', 'SETDA', 'SEKDA', 'POLRES', 'POLDA', 'POLSEK'	**Government**

Example:

Sender	Sender Category	Beneficiary	Beneficiary Category
INDO PRIMA SEMESTA	Business Entity	PT. UNITED FAMILY FOOD	Business Entity
RPKBUNP. SPAN BNI	Government	CV. TANJUNG AGUNG	Business Entity
TRI AMALIA	Others	PT. MAJU MOBILINDO	Business Entity

METHODOLOGY – DATA EXTRACTION

Fund Transfer of SKNBI

- Sender: Business Entity; Recipient: Business Entity → yes → **Business**
- no → Description: NULL, only number (123187937) or punctuation (-,/.:) → yes → **NULL**
- no → Description: Not related to consumption/income → yes → **OTHERS**
- no → Recipient: others; Description: Related to Income → yes → **HOUSEHOLD INCOME**
- no → Sender: others; Description: Related to Consumption → yes → **CONSUMPTION**
- no → **OTHERS**

We use rule-based (keyword) approach for classifying transaction purpose (consumption, income, business)

Household Income Keywords	'gaji', 'honor', 'upah', 'payroll', 'salary', 'remunerasi', 'insentif', 'wage', 'sales', 'pensiun', 'lembur', 'overtime', 'dividen', 'kompensasi', 'bagi hasil', 'bonus', 'claim', 'klaim', 'payoneer', 'komisi', 'tukin', 'uang makan'
Household Consumption Keywords	'pembayaran', 'konsumsi', 'belanja', 'pelunasan', 'kos', 'cicilan', 'dp', 'payment', 'dana', 'pelunasan', 'setoran', 'angsuran', 'jasa', 'invoice', 'listrik', 'service', 'sewa rumah', 'kasbon', 'catering', 'premi', 'asuransi', 'umroh', 'haji', 'tiket', 'spp', 'donasi', 'air ', 'kuliah', 'obat', dsb.

Keywords that are not related to consumption/income
'retur', 'return', 'tabungan', 'refund', 'saving', 'reimburse', 'pemindahbukuan', 'nabung', 'tsa', 'span', 'pemerintah', 'pajak', 'sp2d', 'sppd', 'pendes', 'dana bos'

RESULT & ANALYSIS

In this study, we use F1-score as an evaluation metric. The evaluation was carried out on ±4.000 transactions (random sampling) during the period of 2018 to 2021. **The evaluation results show us a good value of the overall F1-score (84,5%).**

CONFUSION MATRIX EVALUATION RESULT

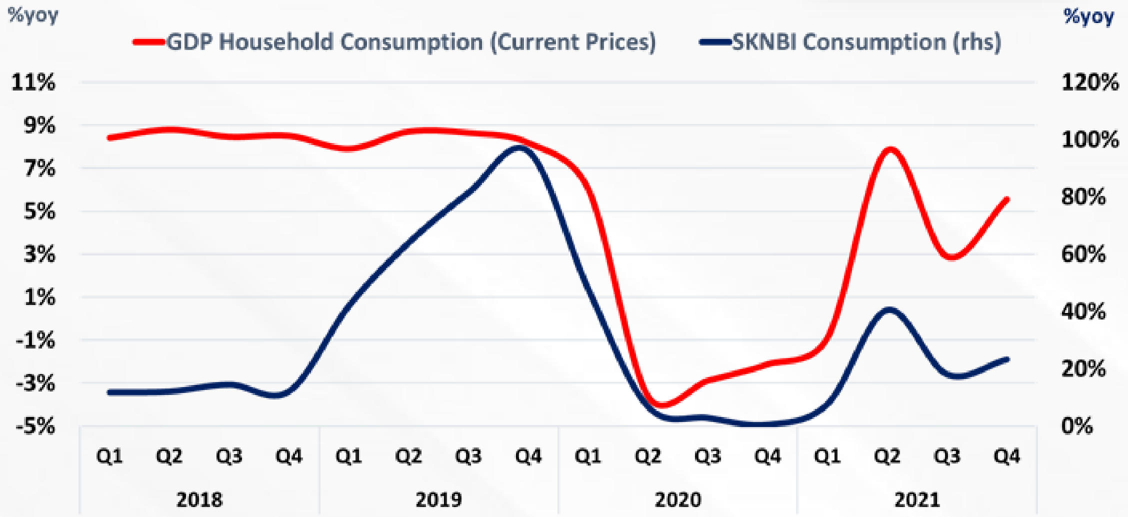
		PREDICTION			
	Purpose of Transaction	**Consumption**	**Income**	**Business**	**Others**
ACTUAL	**Consumption**	154	0	0	38
	Income	17	569	4	76
	Business	42	0	1.078	240
	Others	33	7	12	1.935

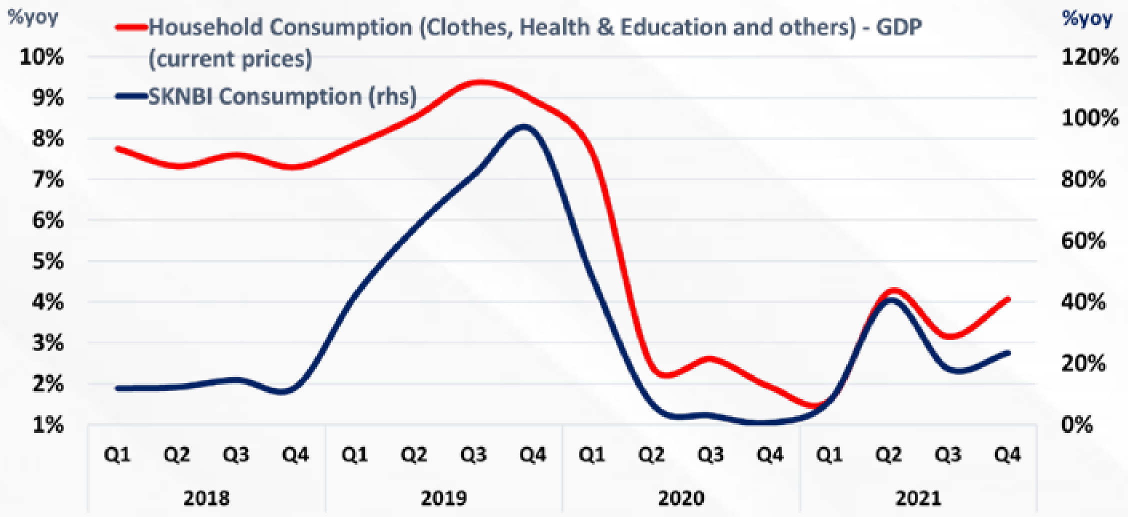
Purpose of Transaction	*Recall*	*Precision*	F1-*score*	F1-*score* (average)	Accuracy
Consumption	80,2%	62,6%	70,3%	84,5%	88,3%
Income	85,4%	98,8%	91,6%		
Business	75,6%	98,2%	85,4%		
Others	97,4%	84,5%	90,5%		

RESULT & ANALYSIS

The validation results show a high correlation between the two indicators since quarter 1-2019, including during the Covid-19 pandemic.

Indicators	Correlation of SKNBI Consumption Growth Rate with GDP		
	Q1-2018 s.d. Q4-2021	Q1-2019 s.d. Q4-2021	Q1-2020 s.d. Q4-2021
Total of Household Consumption	55,2%	85,2%	91,2%
Clothes, Health & Education, Restaurant & Hotel, and others	61,3%	85,2%	90,4%
Clothes, Health & Education and others	67,7%	93,3%	90,3%
Health & Education	65,9%	83,4%	60%
Restaurant & Hotel	50,8%	69,2%	80,2%





CONCLUSION

Conclusion

1. We have **proposed a new approach** in utilizing high-frequency payment system transaction data, **using text mining methodology through a rule-based model**, to construct a proxy indicator for household consumption. **Based on the evaluation of the model, the average F1-score of the model is 84,5%.**
2. Our consumption indicator can be **generated from payment system data more quickly** compared to household consumption indicators in GDP publications. The validation results show a **high correlation** between our consumption indicator from payment system and publication of GDP data, which indicates that the indicator from payment system can be used as **a proxy for household consumption indicators**.

Future Works

1. **Improving the methodology for classifying customer categories and transaction purposes**, including the use of machine learning algorithms.
2. **Utilizing Bank Indonesia's Fast Payment (BI-FAST) transaction data** (implemented since the end of 2021) as additional data source in constructing consumption indicators from retail payment system.
3. Using consumption indicators from payment system, e.g. funds transfer from SKNBI customers or payment transactions via cards, with other macroeconomic variables, to **construct the nowcasting model of household consumption**.

BANK INDONESIA
BANK SENTRAL REPUBLIK INDONESIA

THANK YOU
TERIMA KASIH

Alvin Andhika Zulen, Mohammad Khoyrul Hidayat, Muhammad Abdul Jabbar, Renardi Ardiya Bimantoro
Statistics Department – Bank Indonesia
Email: alvin_az@bi.go.id, moh_khoyrul@bi.go.id, muhammad_abdul@bi.go.id, renardi_ardiya@bi.go.id