

IndoMPF: Dataset and Tabular-Graph Numerical Reasoning Baseline Model for Generating Notes to Government Financial Statements

REZA DARMAWAN

33462135

Master of Data Science
Faculty of Information Technology

DERRY WIJAYA

Supervisor
Faculty of Information Technology

BACKGROUND



Notes to financial statements accompanies a financial statement to enhance its understandability and to avoid misunderstanding.



The current development of Language Model that allows for various text-generation tasks brings the potential of automating Notes to Financial Statement generation.

INHERENT COMPLEXITY

Construction of notes to financial statements contain complexity, beyond accurate summarization and explanation of certain facts from source table.

B.2.1.2. Natural Resources Revenue	
A	Revenues from Natural Resources, Net
Revenue from Natural Resources, Net	Rp104,203,031,073,614 or reached 113.13 percent of the amount of the total central government expenditure in FY 2023. The total revenues in FY 2023 were Rp14,937,859,197,170 or 4.42 percent of the total central government expenditure in FY 2023. The total realization of central government expenditure in FY 2023 was Rp1,152,229,317,443,116. The total realization of expenditure from state treasurer must be 100 percent of the total expenditure.
Revenue from Natural Resources, Net	Rp268,770,096,278,784 or 100 percent of the total central government expenditure in FY 2023.
Revenue from Natural Resources, Net	Rp104,203,031,073,614 or reached 113.13 percent of the amount of the total central government expenditure in FY 2023. The total revenues in FY 2023 were Rp14,937,859,197,170 or 4.42 percent of the total central government expenditure in FY 2023. The total realization of central government expenditure in FY 2023 was Rp1,152,229,317,443,116. The total realization of expenditure from state treasurer must be 100 percent of the total expenditure.
Revenue from Natural Resources, Net	Rp268,770,096,278,784 or 100 percent of the total central government expenditure in FY 2023.

Central Government Spending by Organization/Budget Division					
The realization of Central Government Expenditure in FY 2023 according to the Ministry of Finance is Rp1,152,229,317,443,116. The total realization of central government expenditure in FY 2023 was Rp14,937,859,197,170 or 4.42 percent of the total central government expenditure in FY 2023. The total realization of expenditure from state treasurer must be 100 percent of the total expenditure.					
The realization of Central Government Expenditure in FY 2023 according to the Ministry of Finance is Rp1,152,229,317,443,116. The total realization of central government expenditure in FY 2023 was Rp14,937,859,197,170 or 4.42 percent of the total central government expenditure in FY 2023. The total realization of expenditure from state treasurer must be 100 percent of the total expenditure.					
The realization of Central Government Expenditure in FY 2023 according to the Ministry of Finance is Rp1,152,229,317,443,116. The total realization of central government expenditure in FY 2023 was Rp14,937,859,197,170 or 4.42 percent of the total central government expenditure in FY 2023. The total realization of expenditure from state treasurer must be 100 percent of the total expenditure.					
The realization of Central Government Expenditure in FY 2023 according to the Ministry of Finance is Rp1,152,229,317,443,116. The total realization of central government expenditure in FY 2023 was Rp14,937,859,197,170 or 4.42 percent of the total central government expenditure in FY 2023. The total realization of expenditure from state treasurer must be 100 percent of the total expenditure.					

Task A: Summarizing a row in the table

Task B: Explanation based on the value from other rows

“What should be included?”

Understanding contextual relationships among rows to execute comparative reasoning

“What should be excluded?”

Exclusion of certain rows when performing numerical operation.

OBJECTIVE

Construction of table question answering dataset with numerical reasoning with specialized domain in macroeconomic and public finance

Fine tuning of embedding model using constructed domain-specific dataset for table retrieval task.

Development of baseline model for table question-answering with numerical reasoning

DATASET CONSTRUCTION

Dataset Sources



- PDF tables
- Definitive file tables
- Monthly Econ Report
- Financial Statement

Table Extraction

Table Dataset with Context Paragraph

Link table-sourced reference text to its corresponding data in dataset

General Information

Table Sourced

Period-Specific

Unusable

Reference Text Extraction and Categorization

Reference Text Categorization

Reference Text Categories

Human Expert Evaluators (2)

Education background: Master of Banking and Finance and Accounting Major graduate

Professional experience: government internal auditor

1. Question and Numerical Reasoning Annotation

Generate Question and Reasoning Annotation using LLM

Generate Additional Question and Reasoning Annotation as enrichment

Human expert evaluation

Automatic error detection & correction using audit module

2. Quality Control

Human Expert Evaluators (2)

Education background: Master of Banking and Finance and Accounting Major graduate

Professional experience: government internal auditor

Table Dataset with Question and Reasoning Annotation

Reference Text Dataset with Corresponding Table & Question

Available Operators in the IndoMPF Dataset

SINGLE OPERATION

retrieve (1 operand)

retrieve certain data from the table

addition (2 operands)

The total between two operands.

subtraction (2 operands)

Subtraction, the order of operands represent the minuend and the subtrahend.

multiplication (2 operands)

Multiplication, with the first operand become the multiplier and the second the denominator.

division (2 operands)

Division, with the first operand become the dividend and the second the divisor.

greater (2 operands)

Compare two values to find the greater, yields true or false.

smaller (2 operands)

Compare two values to find the smaller, yields true or false.

average (1 operand)

To calculate the average of values given.

count (2 operands)

To identify the number of values given.

covert (2 operands)

To convert the value of elements value that satisfy criteria to certain predefined value. The first operand is the value that will be compared with the other value.

filter (2 operands)

To filter the elements that satisfy certain predefined value. The first operand is the filter criteria and the second will be compared with the other value.

maximum (2 operands)

To identify the number of maximum of values given. The first operand is a constant and the second is the value to be taken. It can return more than one value, depends on 'n' of the first operand.

minimum (2 operands)

To identify the minimum of values given. Only returns one value.

trace (1 operand)

Get the value of other column in the same row. The first operand is the index of the row to be traced. The second operand is the 'source' of the operand is 'constant'. The second operand is the value to be traced.

divide (2 operands)

Divide, with the first operand become the dividend and the second the divisor.

sum (2 operands)

To identify the minimum of values given. Only returns one value.

average (2 operands)

To calculate the average of values given.

count (2 operands)

To identify the number of values given.

covert (2 operands)

To convert the value of elements value that satisfy criteria to certain predefined value. The first operand is the value that will be compared with the other value.

filter (2 operands)

To filter the elements that satisfy certain predefined value. The first operand is the filter criteria and the second will be compared with the other value.

maximum (2 operands)

To identify the number of maximum of values given. The first operand is a constant and the second is the value to be taken. It can return more than one value, depends on 'n' of the first operand.

minimum (2 operands)

To identify the minimum of values given. Only returns one value.

divide (2 operands)

Divide, with the first operand become the dividend and the second the divisor.

sum (2 operands)

To identify the minimum of values given. Only returns one value.

average (2 operands)

To calculate the average of values given.

count (2 operands)

To identify the number of values given.

covert (2 operands)

To convert the value of elements value that satisfy criteria to certain predefined value. The first operand is the value that will be compared with the other value.

filter (2 operands)

To filter the elements that satisfy certain predefined value. The first operand is the filter criteria and the second will be compared with the other value.

maximum (2 operands)

To identify the number of maximum of values given. The first operand is a constant and the second is the value to be taken. It can return more than one value, depends on 'n' of the first operand.

minimum (2 operands)

To identify the minimum of values given. Only returns one value.

divide (2 operands)

Divide, with the first operand become the dividend and the second the divisor.

sum (2 operands)

To identify the minimum of values given. Only returns one value.

average (2 operands)

To calculate the average of values given.

count (2 operands)

To identify the number of values given.

covert (2 operands)

To convert the value of elements value that satisfy criteria to certain predefined value. The first operand is the value that will be compared with the other value.