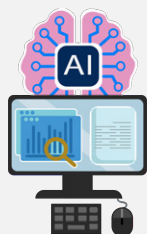


# Use of Artificial Intelligence to Extract Unstructured Data from Financial Statements

by Lynn Zhang  
Business Statistics Division  
Singapore Department of Statistics

## Introduction



The Singapore Department of Statistics (DOS) adopts an Administrative-and-Big-Data-First strategy in the compilation of official statistics. Under this approach, administrative and other non-survey data sources are utilised where possible before conducting statistical surveys. This reduces respondent burden, lowers data collection costs, and supports the production of more timely and comprehensive statistics.

With technological advancements, DOS is exploring the use of Artificial Intelligence (AI) and Machine Learning (ML) to extract information contained in unstructured administrative data such as financial statements.

This article highlights DOS's development of Integrated Business Survey System (IBSS)Pro-AI, an AI-enabled system that automates the extraction of information from firms' financial statements to support the compilation of business statistics.

## Extracting Valuable Insights from Financial Statements

Business statistics compiled by DOS rely extensively on administrative data, with a key source being the financial statements that firms submit to the Accounting and Corporate Regulatory Authority (ACRA).

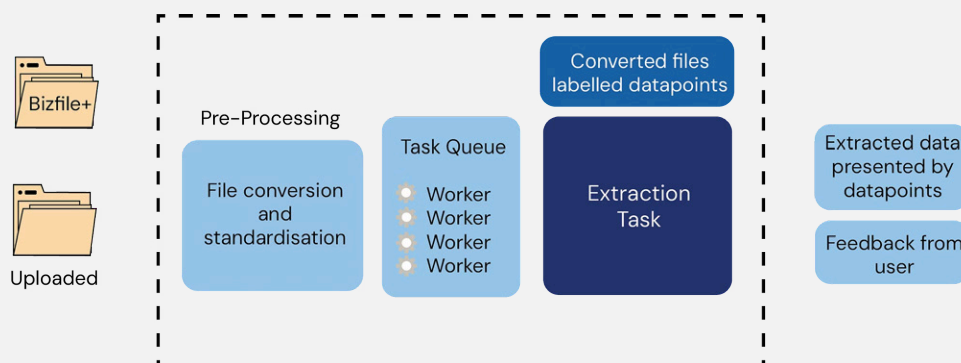
Since 2014, DOS has utilised financial statements filed in the eXtensible Business Reporting Language (XBRL) format — a structured and machine-readable way of reporting financial information. However, only selected variables are mandatory under XBRL reporting requirements, in which other detailed financial items required for statistical compilation are only available in the complete financial statements submitted in PDF or Rich Text Format.

These documents contain valuable information but are largely unstructured, making automated processing difficult. Prior to the adoption of AI-enabled solutions, case officers manually reviewed each financial statement and extracted the required information. Depending on the complexity of the financial statement, this process could take up to one hour per case.

## Developing IBSS Pro-AI

In 2024, DOS developed IBSS Pro-AI to improve operational efficiency and further leverage administrative data. The AI-enabled system automates the extraction of information from financial statements, reduces the manual processing, and enhances the timeliness and comprehensiveness of business statistics. The extraction workflow in IBSS Pro-AI is illustrated in Figure 1.

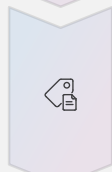
Figure 1: IBSS Pro-AI Extraction Workflow



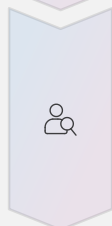
The workflow consists of three main stages.



Firstly, financial statements are ingested into the system and pre-processed daily for new files to be incorporated in a timely manner. During this stage, the documents are converted into a standardised machine-readable format using Optical Character Recognition, preparing them for subsequent analysis.



Secondly, IBSS Pro-AI applies a combination of rule-based extraction techniques and Natural Language Processing (NLP) to identify and retrieve relevant information from financial statements. Rule-based techniques leverage accounting conventions, semantic anchors and table layouts to efficiently identify variables. NLP complements this by recognising synonyms and variations, such as identifying 'Turnover' as 'Revenue', thereby improving the flexibility and accuracy of the extraction. Named Entity Recognition further enhance the process by recognising and labelling text such as dates, organisations, and currency.



Finally, the system incorporates a human-in-the-loop feedback mechanism where case officers validate the extracted information against the original financial statements and make corrections where necessary. These corrections are looped back into the system to refine extraction rules and improve extraction accuracy.

Validation rules are also integrated to ensure extracted values conform to defined guidelines such as ensuring variables like total assets or revenue are non-negative values. The system also provides export functionality to facilitate downstream statistical processing.

## Enhancing Statistical Compilation with IBSS Pro-AI

IBSS Pro-AI has substantially improved the efficiency of processing firms' financial statements.

The system can extract more than 100 qualitative and quantitative variables including principal activity, detailed income, expenditure and fixed asset items commonly collected through DOS's Annual Industry Survey (Figure 2). By integrating extracted data into the statistical compilation process, IBSS Pro-AI has helped to reduce respondent burden and data collection costs.

**Figure 2: Example of the Extracted Data from Financial Statement in IBSS Pro-AI**

### Firm's Financial Statement

|                                                                                                                                                                                                                                                                 |  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| These notes form an integral part of and should be read in conjunction with the accompanying financial statements.                                                                                                                                              |  |
| <b>1. CORPORATE INFORMATION</b>                                                                                                                                                                                                                                 |  |
| Abcdef Ghi Pte. Ltd. is incorporated in Singapore with its principal place of business and registered office located at 13XXXXXXX Avenue 1, 01-1234 Singapore 123456.                                                                                           |  |
| The principal activities of the Company are those of general trading and the beauty, health spa consultants and operators. There have been no significant changes in the nature of activities during the financial year.                                        |  |
| The immediate and ultimate holding company is Abcdef Subsidiary Ltd, incorporated in Singapore.                                                                                                                                                                 |  |
| The financial statements of the company were authorised for issue in accordance with a resolution of the director on 18 June 2026.                                                                                                                              |  |
| <b>2. MATERIAL ACCOUNTING POLICY INFORMATION</b>                                                                                                                                                                                                                |  |
| <b>2.1 Basis of preparation</b>                                                                                                                                                                                                                                 |  |
| The financial statements of the Company have been drawn up in accordance with Financial Reporting Standards in Singapore (FRSs). The financial statements have been prepared on the historical cost basis except as disclosed in the accounting policies below. |  |
| The financial statements are presented in Singapore Dollars (\$), which is the Company's functional currency. All financial information presented in Singapore Dollars has been rounded to the nearest one dollar.                                              |  |

### Extracted Data in IBSS Pro-AI

|                                                         | Extracted                                                                                                                |
|---------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| <b>Company's Financial Year-end ()</b><br>Generic Level | 2025-12-31 -%                                                                                                            |
| <b>Principal Activity ()</b><br>Generic Level           | principal activities of the Company are those of general trading and the beauty, health spa consultants and operators -% |
| <b>Currency Reported ()</b><br>Generic Level            | SGD -%                                                                                                                   |
| <b>Units Reported ()</b><br>Generic Level               | Actual -%                                                                                                                |
| <b>Current period start date ()</b><br>Generic Level    | 2025-01-01 -%                                                                                                            |

Operational efficiency has significantly improved. Previously, case officers typically spent up to one hour manually extracting information from a financial statement. With IBSS Pro-AI, officers can now validate approximately two to three financial statements within the same timeframe, allowing resources to be deployed more effectively while maintaining data quality.

## Conclusion

By combining administrative data with AI and ML technologies, DOS has transformed a manual data extraction process into a more efficient and scalable workflow.

Looking ahead, DOS will continue the adoption of AI, including integrating Large Language Models to improve contextual understanding of complex financial statements to strengthen extraction accuracy of such document, and enhance the system's ability to interpret related information across different sections of a financial statement.