Implementation of Singapore Standard Industrial Classification 2025 in the Statistical Business Register

By **Irmawati Khusaini** and **Hammadah Hassan**, Business Statistics Division
and **Jeremy Lim**, Research Division
Singapore Department of Statistics

Introduction

The Singapore Standard Industrial Classification (SSIC) is the national standard for classifying economic activities undertaken by economic units. It is widely used in the Singapore Censuses of Population, household and establishment surveys, as well as administrative and statistical databases. The SSIC is reviewed and updated regularly to reflect developments in the Singapore economy, particularly the emergence of new economic activities, and to align with changes in international standards.

The SSIC 2025 is the latest edition of the SSIC and adopts the basic framework of the latest International Standard Industrial Classification of All Economic Activities Revision 5 (ISIC Rev. 5) developed by the United Nations Statistics Division. It has a hierarchical structure with a 5-digit coding system, which progressively shows finer details from the highest numeric level of aggregation (Division) to the lowest level (Sub-class). This provides the flexibility to tabulate, analyse and publish data according to the level of detail required. An illustration of the hierarchical nature of the SSIC is shown in Table 1.

Table 1: Example of the SSIC Classification Structure

Classification Level	Classification Title	Classification Code
Section	Manufacturing	C
Division	Manufacture of Computer, Electronic and Optical Products	26
Group	Manufacture of Electronic Components and Boards	261
Class	Manufacture of Semiconductor Devices	2611
Sub-class	Semiconductor Wafer Fabrication	26112

This article [1] presents the implementation of SSIC 2025 in the Statistical Business Register (SBR), which serves as the foundation for compiling business and economic statistics. It discusses the strategy and approach in updating the statistical classification of economic units in the SBR and the Singapore Department of Statistics' (DOS) experience in leveraging artificial intelligence techniques to the statistical classification.

Implementation of SSIC 2025 in the SBR

The SBR is used as the foundational statistical database for business and economic data collection and compilation. It is primarily updated using administrative data and supplemented with statistical survey returns from DOS and Research & Statistics Units (RSUs) in government ministries and statutory boards.

The SSIC is a key indicator for all statistical units in the SBR. It is primarily sourced from administrative data from the Accounting and Corporate Regulatory Authority (ACRA) and other Unique Entity Number (UEN) issuance agencies, and further supplemented by additional information received from statistical survey returns from DOS and RSUs.

In the implementation of SSIC 2025 in the SBR, all firm-level SSIC information were recoded from SSIC 2020 to SSIC 2025. Detailed two-way correspondence tables linking the SSIC 2025 and SSIC 2020 were created to facilitate this recoding process.

[1] Adapted from the paper presented at the **29th Meeting of Wiesbaden Group of Business Registers**.

There are four types of correspondence between SSIC 2020 and SSIC 2025:

1-to-1:

Direct **one-to-one** correspondence between SSIC 2020 and SSIC 2025

Many-to-1:

Multiple SSIC 2020 sub-classes are recoded into **one** SSIC 2025 sub-class, due to a collapsed classification structure

1-to-Many:

One SSIC 2020 sub-class corresponds to **multiple** SSIC 2025 sub-classes to provide greater disaggregation

Many-to-Many:

Multiple SSIC 2020 sub-classes correspond to **multiple** SSIC 2025 sub-classes, due to changes in classification structure

Recoding of SSIC sub-classes in the 1-to-1 and Many-to-1 correspondence types is straightforward and can be performed with certainty. However, recoding of SSIC sub-classes in the 1-to-Many and Many-to-Many correspondence types is more complex as firms can be recoded to more than one SSIC 2025 sub-class. Hence, more effort was required to assign the most suitable SSIC 2025 sub-class during the recoding of these firms. Table 2 presents the SSIC sub-classes and share of enterprises affected by the type of correspondence.

Table 2: Number of SSIC Sub-classes and Share of Enterprises Affected, by Type of Correspondence

Type of SSIC Correspondence	Number of SSIC 2020 Sub-classes Affected	Proportion of Enterprises Affected
1-to-1	865	83.6%
Many-to-1	50	1.1%
1-to-Many	3	0.3%
Many-to-Many	96	15.0%
Total	**1,014**	**100%**

Implementation Strategy

A multi-approach implementation strategy was developed for the SSIC recoding in the SBR:

Adopt a mixture of automated and manual recoding to account for the complexity and data impact

Leverage readily available administrative data (e.g., licensing information) and survey returns

Develop Machine Learning (ML)-based classification tool, the SSIC Remapper

Firms in the SBR are categorised into four groups in the SSIC recoding process, factoring in the coding logic, data availability, economic significance, and complexity of the firms (Figure 1).

Group 1: Automated recoding

Applied to 84.7% of the firms in the SBR whose SSIC 2020 sub-classes fell in the 1-to-1 or Many-to-1 correspondence types as the recoding is clear-cut. These sub-classes were automatically recoded to their corresponding SSIC 2025 sub-classes based on the correspondence table.

Group 2: Manual recoding

Among firms that could not be automatically recoded, large and complex enterprises were identified and manually recoded given the complexity in classification and significant impact on data. A suitable SSIC 2025 sub-class will be assigned after a comprehensive review. About 1.6% of firms fall into this category.

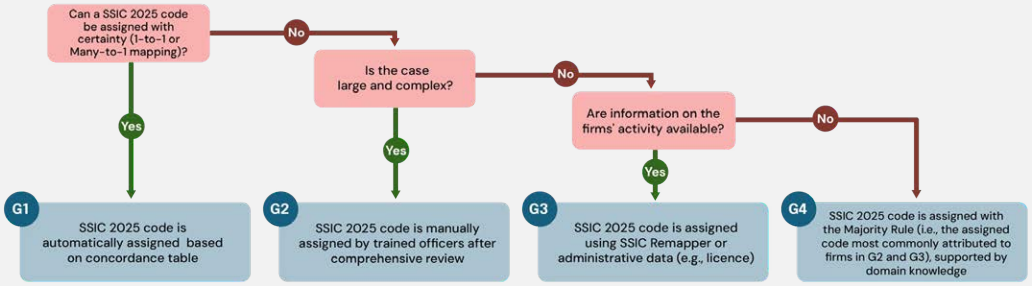
Group 3: Administrative data and ML-based recoding

In the 1-to-Many and Many-to-Many correspondence types, non-large and non-complex firms with available economic activity data were recoded with the SSIC Remapper and supplemented by administrative data. About 0.6% of firms fall into this category.

Group 4: Majority-rule based recoding

The remaining firms were mainly those that could not be recoded with certainty and had insufficient information about their economic activities. These firms were recoded with the majority rule approach, assigning the most frequently attributed SSIC 2025 sub-class to the SSIC 2020 sub-class. This approach was determined with the SSIC recoding outcomes in Groups 2 and 3, supported by domain knowledge. About 13.1% of firms fall into this category.

Figure 1: Process Flow for SSIC 2025 Recoding

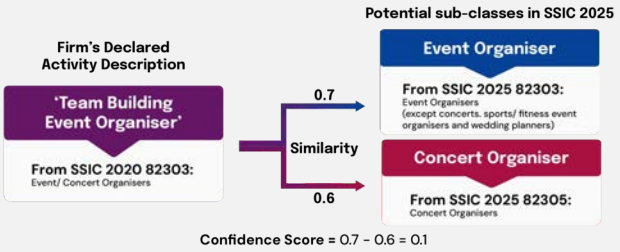


ML-Based Recoding

Mapping codes from different SSIC editions is a complex and time-consuming task, especially for those in 1-to-Many or Many-to-Many correspondence types. DOS developed the 'SSIC Remapper' to automate and streamline this process with ML techniques. The SSIC Remapper narrows down the most suitable SSIC sub-class that are closest to the firms' declared activity descriptions.

Figure 2 illustrates how the SSIC Remapper works. It first converts the firms' activity descriptions and definitions of potential SSIC 2025 sub-classes into sentence embeddings [2], representing the semantic meaning of texts into numerical form. Next, the semantic similarities between the firm's activity description and corresponding definitions of the potential SSIC 2025 sub-classes are calculated. The sub-class with the highest semantic similarity becomes the SSIC Remapper's top recommendation. Finally, a confidence score is computed by comparing the semantic similarity of the top and second-best recommendation. If the semantic similarity between both recommendations is small, a low confidence score will be produced, flagging for human review and further assessment.

Figure 2: How the ML-Based SSIC Remapper is Used for SSIC 2025 Recoding



Double-Coding and Stakeholder Engagement

Changes in statistical classifications may impact data coherence across the years. Best practice requires data compilers to backcast historical data compiled under the previous classification and recompile it based on the new classification. This enables data users to compare data across different time periods in their analyses.

Double-coding [3] is an important prerequisite for backcasting economic statistics when statistical classifications change. DOS implemented SSIC double-coding to provide both SSIC 2020 and SSIC 2025 sub-classes for all firms in the SBR from reference year 2009 onwards. This allows data compilers to conduct detailed data analysis and apply the appropriate methodology in statistical backcasting.

In addition, DOS has been proactively engaging stakeholders to support efforts in the SSIC 2025 implementation. Related publications and resources on SSIC 2025, such as detailed definitions, alphabetical index, and correspondence tables, are readily available on the **SingStat Website**. DOS has conducted various knowledge-sharing and training sessions on SSIC 2025 and backcasting methodology for government agencies. Furthermore, DOS provided customised data advisory services when requested.

Conclusion

DOS has successfully implemented the SSIC 2025 recoding for all firms in the SBR through a multi-approach implementation strategy. The experience demonstrates the benefits of adopting a hybrid approach, combining technological capabilities with specialist insights to maintain high standards whilst improving efficiency. Particularly, the ML-based classification tool has enhanced both operating efficiency and data accuracy in statistical classification. Human expertise and domain knowledge continue to play a crucial role in developing statistical algorithms and automation tools, as well as improving data quality through contextual validation.

[2] The large language model *bge-base-en-v1.5* is used in sentence embeddings.

[3] Double coding is a transitional methodology where firms are simultaneously classified under both the previous and updated industrial classification systems during the conversion period.

References

Singapore Department of Statistics (2025), **Implementation of Singapore Standard Industrial Classification (2025) in Statistical Business Register**

Singapore Department of Statistics (2025), **Singapore Standard Industrial Classification 2025**

Singapore Department of Statistics (2025), Selecting a Suitable Industry Code in the Singapore Standard Industrial Classification 2025, **Statistics Singapore Newsletter Issue 2, 2025**

Singapore Department of Statistics (2025), **Automating Classification with DOS Intelligent Classification Engine (DICE)**

United Nations, **International Standard Industrial Classification of All Economic Activities (ISIC)**

Eurostat (2025), **Implementation of NACE Rev 2.1**

Eurostat (2025), **Backcasting Manual**