Transforming Official Statistics Through AI/ML: DOS's journey

Stats Café: AI and the future of National Statistical Offices

22 Sep 2025

DEPARTMENT OF STATISTICS SINGAPORE
Empowering You with Trusted Data

Agenda

- Product: 2 key aspects
  - Productivity gains
  - Knowledge creation
- Process: The Hub-and-Spoke Model
- People: Building Capabilities
- Learning Points



Use of AI/ ML in Statistical/ Data Production Processes

Data Value Chain / AI/ ML Method	Collection	Processing & Integration	Compilation & Analysis	Dissemination
Imputation		**Imputation** to improve quality of estimation for services, trade and International Investment Position	**Prediction** of public and private properties' purchase/ resale price by dwelling type and estate **Projection** of near-term estimates using time series modelling with AI/ML	
Classification	**Profiling** Singapore's Internet Economy with Big Data **Classification** to identify overseas affiliates from financial accounts **Classification** of topics of media reports or articles	**AI-enhanced OCR and ML for coding** of household expenditure item descriptions from surveys to S-COICOP DICE **D**OS **I**ntelligent **C**lassification **E**ngine **(DICE)** to recommend SSXC codes based on firms' principal activity description, individuals' job title and main duties	**Classification** of CPI to assign web-scraped prices for clothing and footwear to relevant CPI items based on description	
Automation	**AI to capture data** from unstructured financial statements **Smart web-crawling and scraping** to analyse web content and identify relevant data for extraction	**Automation of data confrontation** between different survey sources to improve data quality and coherence **Survey imputation** for non-response case to improve quality of estimation Machine learning to automate the curation of clean training data for ML **Machine learning toolkit** lowers barrier for officers to quickly implement ML.	**Automation of job portal** to consolidate profiles of potential job applicants into a single view for easier reference **Compilation of CPI** via Python **AI** to perform firm-level analysis on financial statements **AI + LLM** to provide semantic search on unstructured text data and quick sensing of scoped population	**Visual Analytics**: AI features in Tableau allow users to visualise/ analyse without coding **Automated report generation** to minimise preparatory time of various short-term reports  **S**tatistics **AN**d **D**ata **R**etrieval **A**.I. assistant **(SANDRA)**, an LLM AI, to assist the public find data in the SingStat Table Builder

Completed
Ongoing
Upcoming

DEPARTMENT OF STATISTICS SINGAPORE
Empowering You with Trusted Data

As at Sep 2025