

Other applications of Machine Learning

As it progressed, the Machine Learning project was informed about other developments of ML to produce official statistics. In particular, during a series of virtual sessions held in October 2020, several speakers were invited to provide an introduction on ML developments conducted in their statistical organisations. It is important to note that they were not carried out within the ML project. The presentations are shared to further highlight the interest in advancing the use of ML. The list includes abstracts and presentations to papers delivered at the [BigSurv20 - Big Data Meets Survey Science](#) virtual conference held in 2020.

Main statistical process	Development	Data source
Edit and Imputation	Australia - Repairing Big Data sets using KNN	Combination of sources
Edit and Imputation	Australia - Census Occupancy Imputation for Census 2021	Census
Edit and Imputation	Canada - Investigating the use of ML methods in Banff and G-Sam.pdf	All sources
Edit and Imputation	Switzerland - Data validation with machine learning - Plausi++ (document)	Administrative
Edit and Imputation	Switzerland - Improving Data Validation using Machine Learning (presentation)	
Estimation and Analysis	Canada - Deploying Machine Learning Techniques for Crop Yield Prediction	Combination of sources
Estimation and Analysis	Germany - Using administrative data and machine learning to address nonresponse bias in establishment surveys	Administrative
Estimation and Analysis	OECD - Nowcasting Services Trade	Aggregates
Estimation and Analysis	Switzerland - ML_SoSi: Individual trajectories in the social security system	Combination of sources
Data collection	BigSurv20 - Automated double-barreled question classification using machine learning	Survey
Record linkage or matching	Canada - Machine Learning for Record Linkage at Statistics Canada	All sources
Record linkage or matching	BigSurv20 - A novel approach to combine survey and bibliometric data for science policy research	Survey and register
Record linkage or matching	USA BLS - Matching fatal injury records with supervised machine learning	Survey and administrative
Text classification	Belgium Flanders - A better statistic on innovative companies in Flanders using web scraping and machine learning	Web scraped
Text classification	Canada - Automation of Information Extraction from Financial Statements using Graph-Based Techniques	Paper documents
Text classification	European Central Bank - Estimating the institutional sectors of legal entities on a large scale European Central Bank - A supervised learning approach	Register /Administrative
Text classification	France - Coding occupations and coding products: two NLP applications for Official Statistics	Scanner and census
Text classification	Israel - Development of ML model for coding of economic activities and occupations in household surveys	Survey and census
Text classification	Detecting innovative companies via the text on their website (Piet Daas)	
Text classification	OECD - 2020-10 Presentation_OCDE_SDG_FinLab_UNECE_Oct13.pdf	Administrative and metadata
Text classification	Switzerland - Automation of General Classification of Economic Activities coding - NOGAuto	All sources
Text classification	UK - Automated classification of web scraped clothing data in consumer price statistics	Web scraped
Text classification	USA USCB - Comparison of Machine Learning Algorithms to Build a Predictive Model for Classification of Survey Write-in Responses	Write-in responses
Text classification	USA USCB - Shared AI Services Hosting Application	Write-in responses

Text classification	BigSurv20 - A framework for using machine learning to support qualitative data coding	Write-in responses
Estimation and Analysis	BigSurv20 - A novel approach to combine survey and bibliometric data for science policy research	Combination of sources
Text classification	BigSurv20 - A text mining and machine learning platform to classify businesses into NAICS codes	Combination of sources
Sampling	BigSurv20 - Advances in big data and the impact on grid based sampling	Imagery and other
Quality and Ethics	BigSurv20 - Big data: big claims or bigger calamities? Getting real about survey research in the fourth era	Not applicable
Estimation and Analysis	BigSurv20 - Boosted kernel weighting - Using statistical learning to improve inference from nonprobability samples	Not defined
Data collection	BigSurv20 - Can recurrent neural networks code interviewer question-asking behaviors across surveys?	Survey
Image classification	BigSurv20 - Complimenting agricultural surveys in Rwanda: Using deep learning to classify crops from drone imagery	Imagery
Data collection	BigSurv20 - Detecting difficulty in computer-assisted surveys through mouse movement trajectories: A new model for functional data classification	Survey
Text classification	BigSurv20 - Evaluating and improving a text classifier for subpopulations: the case of cyber crime	Survey and Administrative
Text classification	BigSurv20 - Explaining and predicting web survey response with time-varying factors and incidental data	Web survey
Estimation and Analysis	BigSurv20 - How unequal is Croatia? Results from combined survey and administrative tax data	Survey and Administrative
Estimation and Analysis	BigSurv20 - Identifying depression related behaviour in Facebook – an experimental study	Survey and Web
Text classification	BigSurv20 - Improving SHARE translation verification	Text
Estimation and Analysis	BigSurv20 - LEARN4SDGs–A machine learning based poverty mapping exercise in Austria	Survey
Text classification	BigSurv20 - Measuring the validity of open-ended questions: Application of unsupervised learning methods	Survey
Estimation and Analysis	BigSurv20 - Application of data mining methods using demographic survey data: Analyses of attitudes towards gender roles and domestic violence in Turkey	Survey
Text classification	BigSurv20 - Predicting basic human values from digital traces on social media	Web
Estimation and Analysis	BigSurv20 - Subjective wellbeing and the intention to emigrate: A cross-national analysis of 157 countries, 2006-2017	Survey
Several processes	BigSurv20 - Use of big data and machine learning at statistics Canada	Several sources
Text classification	BigSurv20 - Prediction of author's educational background using text mining	Text
Image classification	BigSurv20 - Remote sensing in support of agricultural surveys: use of Sentinel images and UAV-acquired ground truth for crop mapping in Rwanda	Imagery
Quality and Ethics	BigSurv20 - The challenges of legal analysis, between text mining and machine learning	Text
Text classification	BigSurv20 - Training deep learning models with active learning framework to classify "other (please specify)" comments	Survey
Estimation and Analysis	BigSurv20 - Using administrative data and machine learning to address nonresponse bias in establishment surveys	Administrative
Data collection	BigSurv20 - Using generative adversarial active learning to identify poor closed-ended survey responses	Survey
Text classification	BigSurv20 - Using supervised classification for categorizing answers to an open-ended question on panel participation motivation	Survey
Estimation and Analysis	BigSurv20 - What can be predicted from a national health survey? Is cancer one of them?	Survey