BANK ĊENTRALI TA' MALTA
EUROSISTEMA
CENTRAL BANK OF MALTA

CRUMBS MAKE A LOAF USING PRODUCT PRICE DATA TO NOWCAST FOOD INFLATION

Article published in the Research Bulletin 2025, pp. 6-14

CRUMBS MAKE A LOAF USING PRODUCT PRICE DATA TO NOWCAST FOOD INFLATION[1]

Sarah Spiteri, Massimo Giovannini and Umberto Collodel

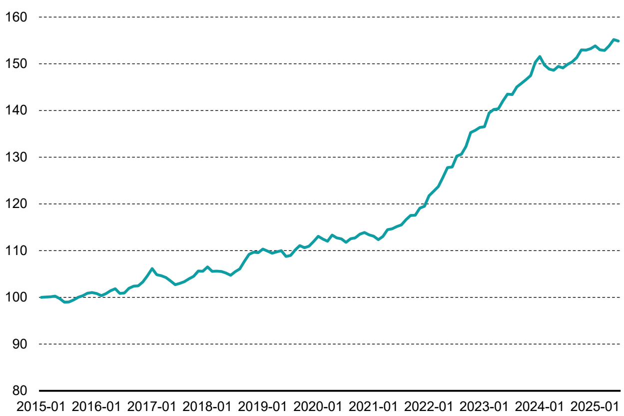
This article explores the use of web-scraped supermarket prices to nowcast food inflation in Malta, where food represents about 16% of the Harmonised Index of Consumer Prices (HICP). We compile a novel dataset of over 2,700 products and more than two million daily prices from supermarkets and corner-shops, classified using string-matching and large language models. Three approaches are assessed: a "naïve" benchmark averaging across products, a minimum distance method aligned with official data, and a machine learning (ML) framework with mixed-frequency regressions. Out-of-sample tests against the Narrow Inflation Projection Exercise (NIPE) show that web-scraped data can enhance forecast accuracy, but predictive accuracy varies by food categories. Results are preliminary but highlight the value of online prices for real-time monitoring.

Introduction

During the COVID-19 pandemic, Malta underwent two structural shifts with direct relevance for the nowcasting of food inflation. First, food inflation accelerated sharply from 2021 (see Chart 1), complicating standard time series approaches for modelling price dynamics due to structural breaks, heightened volatility, and the interaction of global and domestic shocks. Second, lockdowns spurred the digitalisation of food retail, making it feasible to collect high-frequency supermarket prices through web-scraping. These granular data enhance real-time monitoring and nowcasting capacity but raise challenges concerning representativeness, selection bias between online and offline outlets, and consistency with official price statistics.

The food and beverages segment consistently accounts for about 20% of Malta's HICP, with food alone representing roughly 16%. In Malta, food items are primarily purchased from supermarkets and smaller retailers, such as corner-shops. In this light, this study utilises daily online prices from two major supermarket chains and a network of corner-shops, capturing diverse purchasing and pricing behaviours across Malta and Gozo.

Chart 1
FOOD COMPONENT OF MALTESE HICP
(index Level)



Source: Eurostat.
Note: Food component of Maltese HICP in levels from January 2015 to June 2025.

The value of online price data for inflation measurement and forecasting has been widely documented. The Billion Prices Project (Cavallo & Rigobon, 2016) pioneered the use of web-scraped prices to build daily inflation indices across more than 60 countries, showing that such indices closely track – and often anticipate – official Consumer Price Indices. Later work confirmed that online and offline prices are broadly similar, although food retail

[1] The authors would like to thank Governor Edward Scicluna, Deputy Governor Alexander Demarco, Aaron G. Grech, Noel Rapa and attendees to an internal seminar for feedback and comments provided. The views expressed are the authors' own and do not necessarily represent those of the Bank. Any remaining errors are authors' own.

displays larger discrepancies (Cavallo, 2017). Building on this foundation, researchers have incorporated online price indices into econometric and ML models – including MIDAS regressions (Harchaoui et al., 2018), Vector Autoregressions (Aparicio & Urtasun, 2020), and shrinkage methods such as LASSO and Ridge regression (Beck et al., 2023) – consistently improving the timeliness and accuracy of inflation forecasts.

A central limitation of web-scraping is the narrower outlet coverage compared with official statistical samples. While this introduces potential selection bias, the richness and timeliness of item-level data help approximate missing information and partially offset this drawback. Indeed, empirical evidence indicates that even parsimonious models leveraging timely online data can outperform forecasts derived solely from official statistics, notwithstanding their coverage limitations (Macias et al., 2019). Furthermore, the forecaster has no information on which items are monitored by statistical agencies, the weights applied, or the specific calendar days on which prices are collected. Consequently, nowcasting HICP components using item-level prices can be interpreted as approximating this missing information through the exploitation of a rich cross-section of product-level prices. Another layer of difficulty in these studies is that both humans and automated systems face difficulties in accurately categorising products at detailed levels, such as the disaggregated levels of the Classification of Individual Consumption according to Purpose (COICOP) system. Unlike general text classification, product label classification is particularly challenging because names are typically short, incomplete, and structurally different from standard documents, requiring tailored approaches (Yu et al., 2012).

"To our knowledge, this is the first study to apply product-level daily pricing data for inflation nowcasting in Malta, and amongst the first to integrate LLMs into a product classification pipeline"

This article addresses these challenges with three objectives: (i) to construct a high-frequency dataset of retail food prices; (ii) to evaluate its usefulness for nowcasting the components of Malta's food inflation; and (iii) to introduce a novel categorisation pipeline that leverages large language models (LLMs) to assign products to COICOP categories. To our knowledge, this is the first study to apply product-level daily pricing data for inflation nowcasting in Malta, and amongst the first to integrate LLMs into a product classification pipeline.

The remainder of the article is structured as follows. Section 2 describes the data collection and structuring process. Section 3 describes the resulting dataset. Section 4 outlines the nowcasting approaches. Section 5 presents and compares the results. Section 6 concludes.

Data collection and outlet-product panels structuring

We selected three major retailers operating in the Maltese Islands, where two are large-scale supermarket chains, regularly used by a substantial share of households, and a widespread network of corner-shops serving local neighbourhoods. This combination ensures that the dataset captures both dominant retail formats – supermarkets and convenience outlets – which might differ in pricing structures, product turnover, and promotional strategies. Collectively, these retailers provide broad coverage of consumer food and beverage purchases across the islands.[2]

Daily web-scraping routines recorded the complete product assortment from each retailer's online platform starting from January 2024.[3] Prices excluded delivery fees, and manual checks confirmed that online prices correspond to in-store prices. In particular, for chains with more than three outlets, spot checks across locations – including high and low-tourism areas – revealed only minimal and occasional price heterogeneity, primarily attributable to short delays in updating in-store prices.

[2] Indicative estimates based on company revenue disclosures and secondary industry data suggest that the selected outlets correspond to a market share in the range of one-quarter to one-third. These figures are approximate and should be interpreted with caution, given reporting differences and the absence of comprehensive official data at the outlet level.

[3] Data collection failures – caused by website layout changes, server downtime, or network interruptions – were corrected by substituting missing days with data from the most recent successful scrape. This is motivated by the low frequency of price adjustments (see Data Description section) and ensured a full daily time series without gaps.

The raw data collected electronically underwent a multi-step cleaning and validation process. Product names were standardised to remove inconsistencies. Duplicate entries within each retailer's dataset, as well as short-lived products (i.e. those observed for fewer than seven days), were excluded. To preserve product continuity, a label-change detection procedure was applied, relying on a minimum 80% string similarity threshold (based on Levenshtein distance)[4] combined with near-identical pricing. Products displaying implausible day-to-day price changes exceeding 50% were removed as these were likely attributable to recording or scraping errors.

Following data cleaning, products were classified into five-digit COICOP categories and a more granular supplementary food groupings defined by Malta's National Statistics Office (NSO). When retailer-defined categories matched official classifications, metadata mapping was used. Remaining items were matched to already classified products using label similarity (Levenshtein distance with a similarity score greater than 85%). Residual unclassified items were grouped in batches of forty and processed using publicly available LLMs. Throughout this classification, we make sure to limit as much as possible the stochastic nature of the LLMs. In particular, for each product, the model was queried 100 times, and a category was assigned when at least 80% of responses were consistent. Ambiguous cases were escalated to the more powerful publicly available LLMs, and unresolved classifications were manually reviewed.[5]

To construct the indices of the food-related components of HICP, we account for the representativeness of the sample of products by prioritising high-frequency items sold across multiple retailers. A matching procedure was implemented to detect identical products across outlets. Standardised product names were compared pairwise using Levenshtein distance, with matches accepted when differences were less than or equal to 15% of string length, subject to the matching products having consistent metadata.[6] For sparse categories with fewer than ten matched products, additional unmatched (across retailers) items that satisfied the two-month observations per year criterion were incorporated to maintain adequate representation.

The final dataset comprises a consistent, category-mapped panel of products spanning multiple retailers, providing a robust foundation for the construction of representative and comparable consumer price indices over time. Categorisations were evaluated at the four and five-digit COICOP levels, as well as at the more granular level of disaggregated categories provided by the NSO.

To validate our classification procedure, human reviewers were asked to classify a sample of products, including a shared subset of products. Disagreements amongst the reviewers occurred for 17% of micro-category classifications, underscoring inherent ambiguity at fine levels of disaggregation. Importantly, the rate of disagreement among human reviewers was comparable to the misalignment observed between human and LLM-based classifications, with the degree of misalignment between human and LLM-based classifications decreasing as categorisations become more aggregated, as shown in Table 1. Given the resulting levels of accuracy and the inherent degree of ambiguity estimated from the reviewers' exercise, the results of the classification exercise were deemed to be sufficiently accurate to support the nowcasting exercise.

Table 1
PRODUCT CATEGORISATION RESULTS
Percentage shares

COICOP four-digit	COICOP five-digit	NSO micro-categories
95.0%	82.8%	79.0%

Source: Authors' calculations.

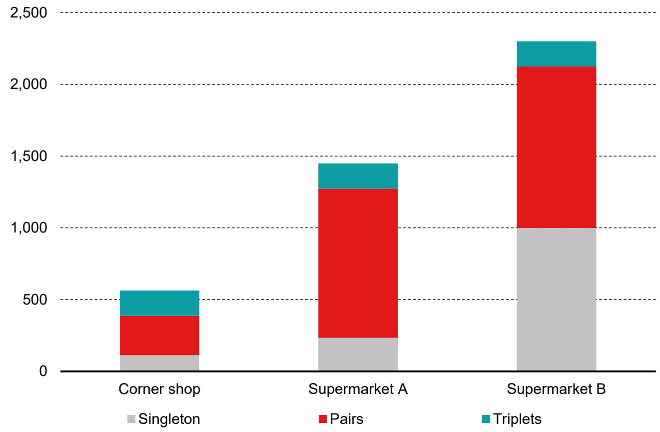
Note: The figures in the table display the fraction of correctly categorised products at the respective level of disaggregation, from the least (left) to the most (right) disaggregated.

4 The Levenshtein distance is a string similarity metric that measures the minimum number of single-character edits – insertions, deletions, or substitutions – required to transform one string into another. It is widely used in natural language processing to assess textual similarity (Levenshtein, 1966).
5 The use of more powerful LLMs was limited to ambiguous cases only, in view of the larger financial costs associated with using these models.
6 Product matches across outlets were required to share the same brand, size, and official category, with price comparability defined as a median price difference not exceeding 25%.

Data description

The final dataset comprises 2,741 distinct food and beverage products distributed across the three sampled outlets, observed from January 2024 to June 2025. This comprehensive collection resulted in 2,237,928 individual price records, providing unprecedented granular coverage of the Maltese food retail market.

Chart 2
PRODUCT COVERAGE ACROSS STORES
(number of distinct products)



Source: Authors' calculations.

Supermarket B maintains the most extensive product assortment with 2,300 items, followed by Supermarket A with 1,449 products, and the corner-shop network with 563 items, as shown by the bars in Chart 2. Product availability across outlets reveals strategic patterns in competitive positioning. The dataset contains 528 products available across all three retail formats, representing core consumer staples such as bread, milk, and basic household items. These universally available products constitute the backbone of food consumption and provide crucial anchor points for cross-retailer price comparisons. The remaining products are distributed as 1,346 singletons (exclusive to one retailer) and 867 pairs (available in two outlets). This distribution pattern suggests that while retailers compete intensively on essential goods, they simultaneously pursue product differentiation strategies to maintain distinct market positions.

Products are distributed across COICOP-4 subgroups with notable concentration in key food categories. The largest segments are sugar, jam, honey, chocolate, and confectionery (599 products, 22% of total) and bread and cereals (572 products, 21% of total). These are followed by vegetables (approximately 15%), dairy products (12%), and meat products (10%). Smaller but important categories include fish and seafood, processed foods, and cooking essentials, with the latter representing the smallest segment at 72 items (3% of total).

Price adjustment behaviour varies significantly across retailers and time horizons. At the weekly frequency, all retailers maintain relatively stable prices, with change probabilities consistently ranging between 1-3%. Monthly price adjustment frequencies reveal the first major divergence in retailer behaviour, with price changes covering between 6% and 12% of products across different retailers sampled. Quarterly analysis amplifies these differences dramatically, with a particular retailer changing prices to 18% of products over three-month periods, with the other two retailers changing prices to 25% and 27% of products**.** Product categories exhibit substantial variation in pricing behaviour. The oils and fats category is the most volatile category, with volatility ranging from 4.2% weekly to 31.3% quarterly. In contrast, other food products n.e.c. (not elsewhere classified) are much less volatile, ranging from 2.4% weekly to 21.4% quarterly.

The distribution of price change magnitudes further reveals contrasting retailer approaches to price management. Two outlets exhibit tightly clustered price adjustments centred around zero, indicating a preference for small, incremental changes that minimise consumer price shocks while allowing for gradual margin adjustments. In stark contrast, the third outlet under consideration displays a markedly asymmetric adjustment pattern characterised by a pronounced concentration of positive price changes.

These heterogeneous patterns across product assortments and pricing behaviours present both opportunities and challenges for nowcasting applications. The diversity in retailer behaviour provides valuable information about different segments of the consumer market, but also complicates the construction of representative indices and the modelling of aggregate price movements.

Indices construction and nowcasting methodologies

The previous section shows that, after cleaning and categorising our products, we obtain a large and heterogeneous set of products that approximates those sampled by the NSO but does not necessarily replicate the full official list and includes non-sampled items. To construct a representative series for the five-digit food components of the HICP, it is therefore necessary to optimise the allocation of product and micro-category weights to closely approximate the official indices. Building on this rationale, we develop three distinct approaches to construct indices and nowcast the food component of the HICP: the first relies on confidential micro-categories price data provided by the NSO, the second adopts a data-driven ML framework independent of such data, and the third is a "naïve" benchmark which uses simple item price averages at the COICOP-5 level. For this purpose, the outlet-product price panel is split into training and testing sets, with May and June 2025 reserved for out-of-sample evaluation.

"We develop three distinct approaches to construct indices and nowcast the food component of the HICP"

The first methodology employs a bottom-up, two-stage minimum distance methodology using confidential monthly price series across the NSO micro-categories. In brief, we first compute monthly price averages across goods and retailers (whenever an item appears in multiple outlets). For each category, we select at most 13[7] items whose price level is sufficiently close to the official series of the micro-category. Second, we minimise the distance between the observed micro-category price level and a weighted average of item prices using constrained least squares, where the weights are the unknown parameters constrained to be non-negative and sum to one. Third, after converting the fitted micro-category price series into growth rates, the same minimum distance procedure is applied to estimate non-negative weights that map NSO micro-categories' inflation to those at the COICOP-5 level. Using the two sets of weights estimated on data up to April 2025, we compute the fitted values for HICP at the COICOP-5 level as well as their out-of-sample nowcast for May and June 2025.

The ML approach also begins with data preprocessing. Price series within each micro-category are cleaned to remove outliers, and for products available in multiple outlets, price series are consolidated by averaging across outlets. To further reduce the dimensionality of the dataset, prices are summarised into micro-category indices using the Jevons index methodology, following Cavallo and Rigobon (2016). The Jevons index, defined as the geometric mean of price relatives, is well-suited to high-frequency web-scraped data due to its robustness to outliers and independence from expenditure weights. Formally, for a set of *n* items, the index at time *t* relative to a base period 0 is given by:

$$J_t = \left(\prod_{i=1}^{n} \frac{p_{i,t}}{p_{i,0}}\right)^{1/n},$$

where $p_{i,t}$ denotes the price of item *i* at time *t*.

Daily Jevons indices are aggregated to weekly frequency and transformed into month-on-month growth rates, which serve as predictors for the monthly five-digit HICP indices. A suite of mixed-frequency and regularised regression models is then applied. Specifically, the modelling set includes: MIDAS (Mixed Data Sampling), which incorporates high-frequency information through distributed lag polynomials; UMIDAS (Unrestricted MIDAS), which flexibly includes individual high-frequency lags; LASSO (Least Absolute Shrinkage and Selection Operator), which performs shrinkage and variable selection using an L_1 penalty; Ridge regression, which applies an L_2 penalty to mitigate multicollinearity; and ElasticNet, which combines L_1 and L_2 penalties to balance sparsity and group selection. Model hyperparameters – such as lag structure, lag polynomial type, and regularisation strength – are optimised via a tailored grid search for each five-digit COICOP category.

While the first methodology applies estimated product and micro-category weights to out-of-sample prices, aggregating results to COICOP-5 predictions, the ML approach generates COICOP-5 nowcasts directly from the

[7] Given the size of the training sample, *T*, of 15 months, to avoid overfitting, we keep at least two degrees of freedom ensuring that T-$N \geq 2$ where *N* is the number of selected products in each category.

trained models, with the model achieving the lowest out-of-sample root mean squared error (RMSE) selected for each category. Finally, predictions from all approaches are aggregated from the five-digit to the four-digit level using official, publicly available weights.

Results

Our nowcasting methodologies were employed to generate both within-sample fits and out-of-sample predictions, shown in Chart 3 and separated by the vertical dashed line. Furthermore, to show the effectiveness of these two approaches in matching actual HICP indices, we compute a "naïve" measure for HICP by calculating simple product average growth rates as per the classification, without undergoing any further optimisation of weights. Furthermore, for brevity, we further aggregate our COICOP-5 indices to COICOP-4.

For the within-sample fit, the "naïve" benchmark performs worst relative to the other two approaches, while the relative performance of the NSO-based and ML methodologies varies across COICOP-4 categories. For example, the ML approach closely approximates the official series for Bread & Cereals, Oils & Fats, Fruit and Vegetables. However, in categories characterised by greater product heterogeneity, such as Oils and Fats and Vegetables, it exhibits signs of overfitting, leading to a marked deterioration in nowcasting accuracy compared with the other approaches, despite strong within-sample performance. By contrast, the NSO-based methodology shows no evidence of overfitting and delivers a consistent within-sample fit, with particularly strong performance in the Milk, Cheese, and Eggs category. The prominence of staple products and branded items in this category appears to facilitate its ability to replicate or closely approximate the provided NSO micro-category aggregated price series.

Chart 3
MONTH-ON-MONTH GROWTH OF FOOD COMPONENT FOUR-DIGIT HICP
(per cent)

Source: Author's calculations.
Note: Month-on-month growth rates for the official indices published by Eurostat (grey), and the predicted values from our ML approach (teal), NSO-based method (red), and the naïve benchmark (lime) by four-digit COICOP categories. The vertical dashed line distinguishes between the in-sample (right) and out-of-sample (left) periods.

To evaluate the performance of our nowcasting approaches, we compare their out-of-sample RMSEs against those from the NIPE. In the NIPE, national central banks generate short-term forecasts of overall HICP inflation and its main components by combining advanced modelling techniques with expert judgement. These forecasts are then aggregated by the European Central Bank to construct the euro-area inflation path. Therefore, this exercise provides a well-established benchmark for assessing the accuracy of our nowcasts.

Table 2 reports the relative performance of our nowcasting approaches, where values below (above) one indicate improved (worsened) accuracy compared with the estimates from the NIPE. It must be stressed that the short training sample (15 months) together with the small set of out-of-sample observations (two months) warrants a cautious interpretation of the results, which should be regarded as preliminary.

Conditional on the above caveat, the results show that at least one of our methods is able to outperform the NIPE across all four-digit COICOP categories, with the exceptions of Milk, Cheese & Eggs and Vegetables. The former poses a particular challenge for nowcasting, as some of its key components are subject to administered pricing, where expert judgement offers a clear advantage in anticipating adjustments. Moreover, the limited sample of products for fresh vegetables substantially affects the quality of our data, thereby limiting the reliability of our nowcasts in the latter category. Consequently, within the short training sample, the NSO-based methodology struggles to identify products with a close resemblance to the official series of micro-categories, whilst the ML approach overfits the training sample with the inclusion of potentially spurious correlations.

The nowcasts generated by our NSO-based and data-driven ML approaches generally outperform those obtained using a simple average across products, with the exceptions of the Bread & Cereals and Vegetables categories. The data-driven ML approach outperformed the NIPE in all but two categories and produced the optimal nowcast in five of the nine categories reported in Table 2. In contrast, the NSO-based approach outperformed the NIPE in three categories but yields the optimal nowcast for the Oils & Fats category. This contrast highlights the trade-off between interpretability and nowcasting performance. While the NSO-based approach offers a high degree of interpretability due to its imposed constraints, its use of constrained optimisation may exclude product-level data that are informative for nowcasting. Consequently, this approach delivers gains where the product pool is large and aligns with officially collected data. The data-driven ML approach, by contrast, is explicitly optimised for nowcasting, achieving greater predictive accuracy at the expense of interpretability. However, its current reliance on a limited training period and the exclusion of official statistics, as incorporated in the NSO-based approach, increases the risk of overfitting. Particularly, the figures in Table 2 highlight the vulnerability of the ML approach for overfitting the Fruits and Vegetables categories due to a limited sample of fresh produce.

Table 2

NOWCASTING PERFORMANCE RELATIVE TO NIPE

RMSE ratios

COICOP-4	Products' mean	NSO-based	ML
Bread and cereals	0.43	1.70	0.74
Meat	0.93	0.80	0.76
Fish and seafood	0.60	0.81	0.50
Milk, cheese and eggs	2.25	3.31	1.77
Oils and fats	0.75	0.62	0.94
Fruit	5.63	4.57	0.85
Vegetables	2.16	2.54	13.48
Sugar, jam, honey, chocolate and confectionery	1.61	1.20	0.92
Food products N.E.C.	1.16	1.37	0.83

Source: Authors’ calculations.

Note: The figures in the table display the ratio of out-of-sample RMSEs of our nowcasting approaches, such that values below (above) one indicate improved (worsened) performance relative to the NIPE.

Conclusion

This study set out to explore the potential of granular web-scraped supermarket prices to improve the timeliness and accuracy of food inflation nowcasting in Malta. By assembling a novel high-frequency dataset covering more than 2,700 distinct food and beverage items across three major retail outlets, we built one of the most detailed datasets to date of pricing dynamics in the Maltese retail food sector. Beyond the descriptive insights into product availability, competition, and pricing behaviour across outlets, this article demonstrated how these data can be systematically harnessed through structured methodologies to produce short-term forecasts of HICP food components.

"These findings align with international evidence that relatively parsimonious models leveraging timely online data can outperform forecasts derived solely from official statistics"

Our results suggest that web-scraped data can deliver meaningful gains in predictive performance relative to established benchmarks, including the NIPE. The ML approach, in particular, consistently outperformed the benchmark in most categories, while the NSO-based minimum distance method provided strong performance in selected groups where official price series are closely aligned with the structure of the scraped dataset. Importantly, these results were obtained despite the inherent limitations of the dataset, including the exclusion of certain outlets and product types, and the relatively short evaluation period. These findings align with international evidence that relatively parsimonious models leveraging timely online data can outperform forecasts derived solely from official statistics.

Several broader lessons emerge from this exercise. First, different methodological approaches present a trade-off between interpretability and predictive accuracy. The NSO-based approach offers parameters that are directly interpretable as non-negative weights, providing transparency and consistency with official statistical practices, but may exclude useful item-level signals. The ML approach, by contrast, is explicitly optimised for prediction and thus yields stronger performance, though at the expense of interpretability and with a higher risk of overfitting. Second, the application of LLMs in the product classification pipeline proved valuable in managing the complexity of categorising thousands of short, often ambiguous product labels, underscoring the potential of LLMs to complement traditional statistical techniques in official data contexts.

Looking ahead, the results presented here should be regarded as preliminary, given the limited sample period and the small number of out-of-sample observations. Future research could extend the training horizon, broaden retail coverage and experiment with ensemble methods that combine interpretable and predictive approaches. Moreover, closer integration with official statistical production processes could enhance both the accuracy and the credibility of web-scraping-based nowcasting tools. In this sense, the contribution of this work is twofold: first, to demonstrate the feasibility and value-added of web-scraped price data for inflation monitoring in Malta, and second, to provide a proof of concept for integrating innovative data sources and ML methods into the toolkit of economic analysis and statistical production.

References

Aparicio, D. and Bertolotto, M. (2020), "Forecasting inflation with online prices", *International Journal of Forecasting, 36*(2), pp. 232-247.

Beck, G. W., Carstensen, K., Menz, J.-O., Schnorrenberger, R., and Wieland, E. (2023), "Nowcasting consumer price inflation using high-frequency scanner data: evidence from Germany", ECB *Working Paper Series* 2930, European Central Bank.

Cavallo, A. (2017), "Are online and offline prices similar? Evidence from large multi-channel retailers", *American Economic Review, 107*(1), pp. 283-303.

Cavallo, A. and Rigobon, R. (2016), "The billion prices project: Using online prices for measurement and research", *Journal of Economic Perspectives, 30*(2), pp. 151-178.

Harchaoui, T. M. and Janssen, R. V. (2018), "How can big data enhance the timeliness of official statistics? The case of the US consumer price index", *International Journal of Forecasting, 34*(2), pp. 225-234.

Levenshtein, V. I. (1966), "Binary codes capable of correcting deletions, insertions, and reversals", *Soviet Physics Doklady, 10*(8), pp. 707-710.

Macias, P. and Stelmasiak, D. (2019), "Food inflation nowcasting with web scraped data", NBP *Working Paper* 302, Narodowy Bank Polski.

Yu, H-F., Ho, C-H., Arunachalam, P., Somaiya, M. & Lin, C-J. (2012), "Product Title Classification versus Text Classification", *Technical Report*, National Taiwan University.