

Nowcasting Inflation: Machine Learning for Automated Product Classification

1. Introduction

Price stability is at the core of the mandate of the European Central Bank (ECB). Accordingly, the ECB has a long tradition of monitoring inflation, forecasting it, and studying its determinants. Fundamentally, however, inflation is determined by the behaviour of millions of price setters, so that the ECB also has a natural interest in studying the dynamics of individual prices. The Internet offers a new possibility to scrape information about the price of goods that are sold online, but also of services such as restaurants and travel services. Never before did we have at our fingertips such a vast amount of information on the behaviour of individual prices. This new and so far unavailable data source has attracted the interest of several areas of the ECB and the ECB is currently investigating the usefulness of web scraped data for various purposes. Web-scraped data on individual product prices can provide invaluable input for monitoring and nowcasting inflation, both of which are core activities of the ECB.

This is not the first time that an organisation uses web scraped online product information. One prominent example is the billion prices project (BPP).¹ While this represents excellent work, it does not meet all the needs by economists at a central bank. For example, the BPP does not provide a very fine granular classification of products, but a central bank is interested in understanding all facets of price determination, which cannot be based on coarse categories, such as food or clothing.

Assigning the right categories to products at a detailed level is one key step to improve our understanding of the dynamics of individual prices and inflation rates within and across these categories. However, it is also a challenging task and has not been addressed by previous work at a larger scale:

- 1) It requires significant expertise to do it correctly
- 2) Millions of products need to be classified - not only once but at a continuous basis. There are not enough experts that can do this in time, especially when taking this to the full European level of 28 countries with 24 official languages

¹ <http://www.thebillionpricesproject.com/>

2. The Challenge

Your task is to develop an intelligent way to automatically assign the right category (according to the [COICOP](#) classification scheme, see below) to supermarket products already scraped from online sources.

COICOP has over 300 (sub)-categories. In the EU there is a harmonised classification up to the five-digit hierarchical level, beyond which each country has discretion on the sub-categories (some EU countries even use up to 10-digit levels). Each category is described textually.

An easy example:

Product: Super Banana (a fresh banana)

COICOP Classification:

01 Food and non-alcoholic Beverages

01.1 Food

01.1.6 Fruit

01.1.6.1 Fresh or chilled fruit

This mapping could be done by hand. However, the EU has 28 countries and 24 official languages. We do not have enough experts to classify millions of online supermarket products, and we expect to receive new products every day. Furthermore, there are no “hard” rules that an algorithm can use to categorize a given product. Usually experts are needed to classify them as there are complex definitions and interpretations.

Some product classifications can be indeed rather complex, take this example:

Product: Frozen Sandwich, Sausage, Egg & Cheese

COICOP Classification:

01 Food and non-alcoholic Beverages

01.1 Food

01.1.9 Food products n.e.c.

01.1.6.1 Ready-made meals

This may be intuitive for a person, but an algorithm could also easily classify it as sausage, cheese, eggs or bread.

Another problem is that little information on very similar products might be available. For example the following products would all be categorized in different categories:

- Megabeer 3.5Alc
- Megabeer without alcohol
- Megabeer Ale

3. The Data

You will have the following data and you should take dataset 1 and dataset 2 into account:



Webscraped_Example
.xlsx

- Dataset 1: Scraped Product Data of two online supermarkets from Germany



Markt guru_Example.xl
sx

- Dataset 2: Product data from supermarket leaflets in Germany

Additionally we provide you the COICOP categories including their description.

Note: The full dataset will be made available to you in the Cloud environment for the Hackathon

We have split the datasets for you in 3 parts:

- Train: This can be used for training your models
- Dev: This can be used to evaluate yourself your trained models and hyperparameter selection
- Test (not available to you): This will be used by the Jury to evaluate your solution

4. Evaluation

There are several evaluation criteria. The one with the largest weight in the evaluation will be about the accuracy of your model. Three other criteria will be about the potential business value that your AI

solution will provide. This is also important, because AI solutions are not only about the models, but how they work in a business process as we have described above. Furthermore, we are also interested in an AI solution that has a manageable cost now and in the future and has a predictable behaviour where the users understand how to interact with it and what they can expect.

Remember what has been stated above: The classification is a key step in an important process.

You will present for evaluation your code and your results to the jury. Please do not prepare a Powerpoint presentation or similar slides – use only your code to show and explain your work. When presenting your work tell the jury why you believe that your solution addresses all those aspects.

You will be evaluated based on the following criteria:

- ☐ Accuracy (60%): According to evaluation function provided to you in your infrastructure based on test dataset that you have not seen
- ☐ Generalisability (15%): According to Jury judgement (between 0 missing to 5 very good)
- ☐ Innovation (15%) : According to Jury judgement (between 0 missing to 5 very good)
- ☐ AI Monitoring (10%) : According to Jury judgement (between 0 missing to 5 very good)

Accuracy

Accuracy is an important aspect of the solution. However, accuracy here is not based on traditional measures, such as precision and recall. Remember that COICOP is a hierarchical classification. While we of course are interested in a solution that classifies a product correctly at all levels, we can accept some errors at the most detailed level of the hierarchy. On the other side, having only the highest level of the classification correct is acceptable, it would though reduce significantly the opportunities for inflation analysis.

An example on how we would evaluate the accuracy of your classification – Please note that the weights on each level in the actual evaluation function might differ (but will be made known to you at the beginning of the Hackathon):

Chiquita Banane – Correct classification as by expert

- (a)** 01 Food and non-alcoholic Beverages
- (b)** 01.1 Food
- (c)** 01.1.6 Fruit
- (d)** 01.1.6.1 Fresh or chilled fruit

Chiquita Banane – As classified by an intelligent classifier

(e) 01 Food and non-alcoholic Beverages (correct: 1/1 point)

(f) 01.1 Food (correct: 1/1 point)

(g) 01.1.6 Fruit (correct: 7/7 point)

(h) 01.1.6.3 Dried fruit and nuts (incorrect: 0/7 point)

This classifier would get $1+1+7+0=9$ points out of the 16 ($1+1+7+7$) maximum points for this classification.

You will have in the cloud environment for the Hackathon the evaluation function, so you can use it on the data made available to you. Please use this function for your evaluation as we will use your provided code to evaluate the results based on the test dataset not made available to you.

Please provide ALWAYS for each product the full category up to level 5 as you will also find in the training data. For example, 01.1.6.3 is acceptable. All predictions that are not provided up to level 5 will get 0 points. For example, 01 or 01.1 or 01.1.6 only would NOT be accepted as correct at all.

Generalisability

This criteria is evaluated based on the expert evaluation by the jury. They will rate this criteria on a scale from 0 to 5 relative to all AI solutions of the Hackathon.

The idea here is that we want to avoid that an intelligent classifier works only very good for one supermarket, but in order to use it for another supermarket you need to invest a significant amount of time to create a lot of new training data. While we expect that some training is always needed, we are interested in reducing our costs and the necessary number of experts to create a large amount of additional training data. Ideally, one could also reuse the trained model with little additional training on supermarket websites with a different language. For example, the classifier is trained on German supermarket data, but can be applied with little training on French supermarket data.

It is very important that you make the point from your perspective why you believe that this criteria is addressed for your case. For this, it would be good to have basic support

Innovation

This criteria is evaluated based on the expert evaluation by the jury. They will rate this criteria on a scale from 0 to 5 relative to all AI solutions of the Hackathon.

Innovation is key to know where the journey goes. While you have developed a very good classifier, we are interested also in how this classifier could be extended or built on in the future. Keep in mind that we may also extend it to more complex classification problems, such as clothes or travel services. While it is

NOT the objective of this Hackathon to provide a classifier for those problems, we are interested in how one can deal with growing complexity.

Innovation can mean, but is not limited to (we like to be surprised with things we have not thought about!):

- Use of additional data sources (images, nutritional content, etc.)
- Combination of several models (e.g. an unsupervised and a supervised approach)
- Using novel ML techniques, such as AutoML or Deeplearning, to avoid the large amount of manual data science activities (e.g. feature selection, model selection, hyperparameter selection)
- Including aspects of transfer learning
- Active Learning

Surprise us!

AI Monitoring

This criteria is evaluated based on the expert evaluation by the jury. They will rate this criteria on a scale from 0 to 5 relative to all AI solutions of the Hackathon.

For us, developing a good model and deploying it is just the beginning of the journey. The AI solution is a critical step of a business process that is executed daily with a large amount of data coming in. It is critical that the process including the AI classification still works after days, months, years. We must be reasonable sure what is going on with the classification or if they are issues to be able to fix them in a timely manner.

This can mean, but is not limited to (we like to be surprised with things we have not thought about!):

- Is the classifier on-track?
- Significant change of product attributes that have not been seen before?
- Certainty of classifications by classifier? Should an expert, rather than the classifier, classify a given product?