Project: Predicting Inflation

Data Products Creation & Deployment

Pablo Cancinos | Joaquín Reyes | María Isabel Rivera

Introduction

According to Banco de Guatemala, nowadays the country is reporting a 3.82% inflation at July2021, which is 0.94pp more than last year at July2020. At this point, it is important to enhance that inflation will always exist in our society, people must learn to live with it. However, if we look at the definition of inflation itself:

> "Inflation is the decline of purchasing power of a given currency over time. A quantitative estimate of the rate at which the decline in purchasing power occurs can be reflected in the increase of an average price level of a basket of selected goods and services in an economy over some period. The rise in the general level of prices, often expressed as a percentage, means that a unit of currency effectively buys less than it did in prior periods."

In other words, inflation makes us feel that we can't afford as much as we could last year, mostly, when our incomes remain static from one year to another. Therefore, companies and customers must be open-eyed when it comes to inflation, because it has a direct impact on their purchasing power.

Therefore, as a team of passionate members about analytics, we are seeking to predict the percentual variance, between next month's inflation value and its last year's value, meaning -12 months, so that we can anticipate increasing inflation or decreasing inflation (deflation) periods. For example, when we predict deflation periods, we may promote in our companies & personal expenses, the perks of deflation, and buy more for less or, save the remaining money that we didn't spend in our regular purchases.

This is how we learn to coexist with inflation behavior, in a way that we can also take the most of it when its behavior moves in our favor.

Dataset

Data source comes from an API Request and it's delivered in json format with the following features:

year = year

period = month

periodName = month label

value = value of basic basket

footnotes = attached features

Following, is an output sample is shown as:

```
{'status': 'REQUEST_SUCCEEDED',
 'responseTime': 230,
 'message': [],
 'Results': {'series': [{'seriesID': 'SUUR0000SA0',
     'data': [{'year': '2021',
       'period': 'M07',
       'periodName': 'July',
       'latest': 'true',
       'value': '153.424',
       'footnotes': [{'code': 'I', 'text': 'Initial'}]},
      {'year': '2021',
       'period': 'M06',
       'periodName': 'June',
       'value': '152.720',
       'footnotes': [{'code': 'U', 'text': 'Interim'}]},
      {'year': '2021',
       'period': 'M05',
       'periodName': 'May',
       'value': '151.405',
       'footnotes': [{'code': 'U', 'text': 'Interim'}]},
      {'year': '2021',
       'period': 'M04',
       'periodName': 'April',
       'value': '150.221',
       'footnotes': [{'code': 'U', 'text': 'Interim'}]}
]}]}}
```

After receiving raw data, there is a process of data preparation, which will be introduced as an input to the machine learning model for predicting next month's inflation percentage delta from the present year with last year's inflation record. Data Preparation process involves creating additional features, such as last month value, last 3 months value, and so on. Following, is a sample of the dataset that will be used to train and test our predicting model:

	series id	year	period	value	footnotes	value-1	value-3	value-6	value-12	var-1	var-3	var-6	var-12	var-1_lag	var-3_lag	var-6_lag	var-12_lag
0	SUUR0000SA0	2021	M06	152.567	Initial	151.253	148.839	146.330	144.847	0.008687	0.025047	0.042623	0.053298	0.007883	0.023377	0.034527	0.050237
1	SUUR0000SA0	2021	M05	151.253	Initial	150.070	147.798	146.205	144.018	0.007883	0.023377	0.034527	0.050237	0.008271	0.021058	0.025671	0.041126
2	SUUR0000SA0	2021	M04	150.070	Initial	148.839	146.975	146.314	144.142	0.008271	0.021058	0.025671	0.041126	0.007043	0.017146	0.017563	0.027092
3	SUUR0000SA0	2021	M03	148.839	Interim	147.798	146.330	146.270	144.913	0.007043	0.017146	0.017563	0.027092	0.005600	0.010896	0.011969	0.017942
4	SUUR0000SA0	2021	M02	147.798	Interim	146.975	146.205	146.050	145.193	0.005600	0.010896	0.011969	0.017942	0.004408	0.004518	0.009430	0.014930

From this data preparation output, our target in the predicting model will be ["value-12"], and other columns will go through a feature selection process, in order to select only the ones that will help predict our target variable.

Model

The model is an implementation of a linear regression, with its variation of a Bayesian Ridge model. We started by obtaining the needed data in order to perform a forecast of our variation of the inflation rate. The data we gathered was the previous monthly inflation rates, from 2012 to 2021. Then we calculated the variation of these values and used them in order to perform a regression to obtain the desired variation on the inflation rate. We used a Bayesian Ridge type of regression because by using its hyperparameters we can create a model that generalizes in a more effective way rather than using a typical linear regression.

Additionally, by using the concepts learnt about explainable AI and the Shap values, we were able to differentiate which variables were the most aligned with target value (the variation of the inflation rate).

Deployment

We published the model to the outside world by using an API builder such as Flask. Flask gave us the needed endpoints in order to perform actions to the model. We save the model as a pickle file and everytime we use it on the API we load it and perform its predictions. This is a best practice in order to deploy any model, because it makes it easier for us to act upon the model. The deployment method consists of publishing the Flask's API to a public server where anyone can get access to it.