

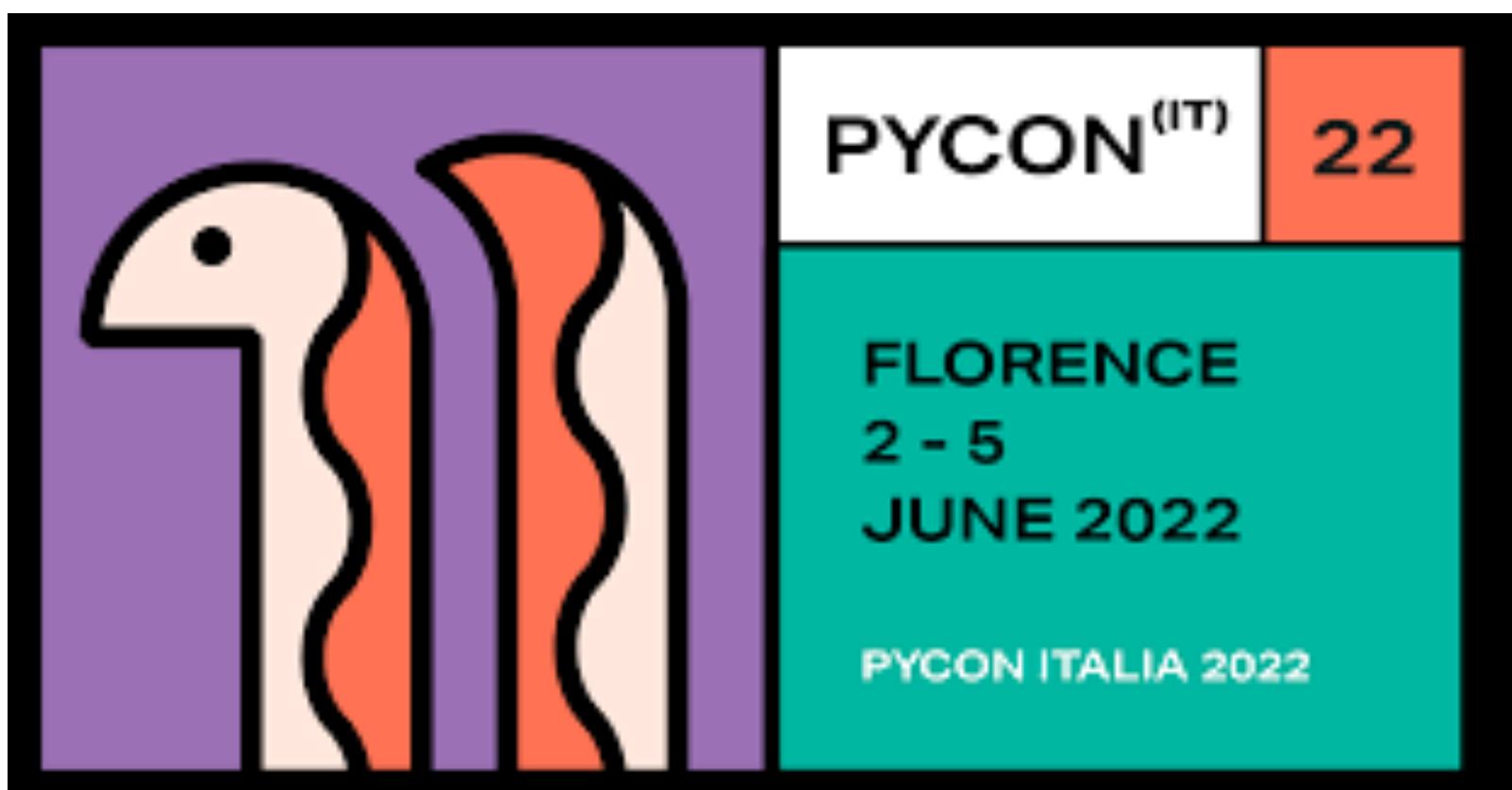
Nowcasting financial crisis with deep learning techniques

Mauro Pelucchi

mauro.pelucchi@gmail.com

<https://github.com/mauropelucchi>

<https://www.linkedin.com/in/mauropelucchi/>



Description

Predict, or project, a financial crisis is important in all the sector: finance, banks, insurance and manufacturing companies take benefits if they can anticipate a crisis situation.

In this tutorial I'll show you how to use python to set up model and analytics to make projections about the next financial situation. I will also try to explain the technical and ethical aspects that we have to consider when we approach these kinds of challenges..

You are provided a git repository (<https://github.com/mauropelucchi/pyconit2022>) with:

- Jupyter notebooks
- Slides

During this training we will use **Google Colab**

Mauro Pelucchi

I am a senior data scientist and big data engineer responsible for the design of the "**Real-Time Labour Market Information System on Skill Requirements**" for



I currently works for



In Emsi Burning glass I work as Head of Data Global Data Science with the goal to explore, design and delivery innovative solutions about Labour Market Data.

In collaboration with the **University of Milano-Bicocca**, I took part in many research projects related to the labour market intelligence systems.

I collaborate with the **University of Milano-Bicocca** as a lecturer at the Master Business Intelligence and Big Data Analytics and with the **University of Bergamo** as a lecturer in Computer Engineering.



Agenda

- Why predict financial crisis?
- Data
- Methodology
- Technical hands-on
- Results
- What next?

Why?



12 Apr 2022

Investors predict US recession in 2023 - here are the facts

The next recession FORTUNE

December 9, 2021

MAY 14 2022

Where the next financial crisis could come from

FINANCIAL TIMES

The
Economist

The next financial crisis

Feb 9th 2022

April 4, 2022,

Bloomberg
Europe Edition ▾

U.S. Recession Seen Most Likely to Start Next Year: Survey

- Bank stocks are seen as underperforming broader market
- Russia seen as offering most-undervalued bonds in survey

The next financial crisis

- Improving investments
- Assessing markets, and new markets
- Supply-chain strategies
- Improving planning
- Adopt better strategies in advance

The next financial crisis

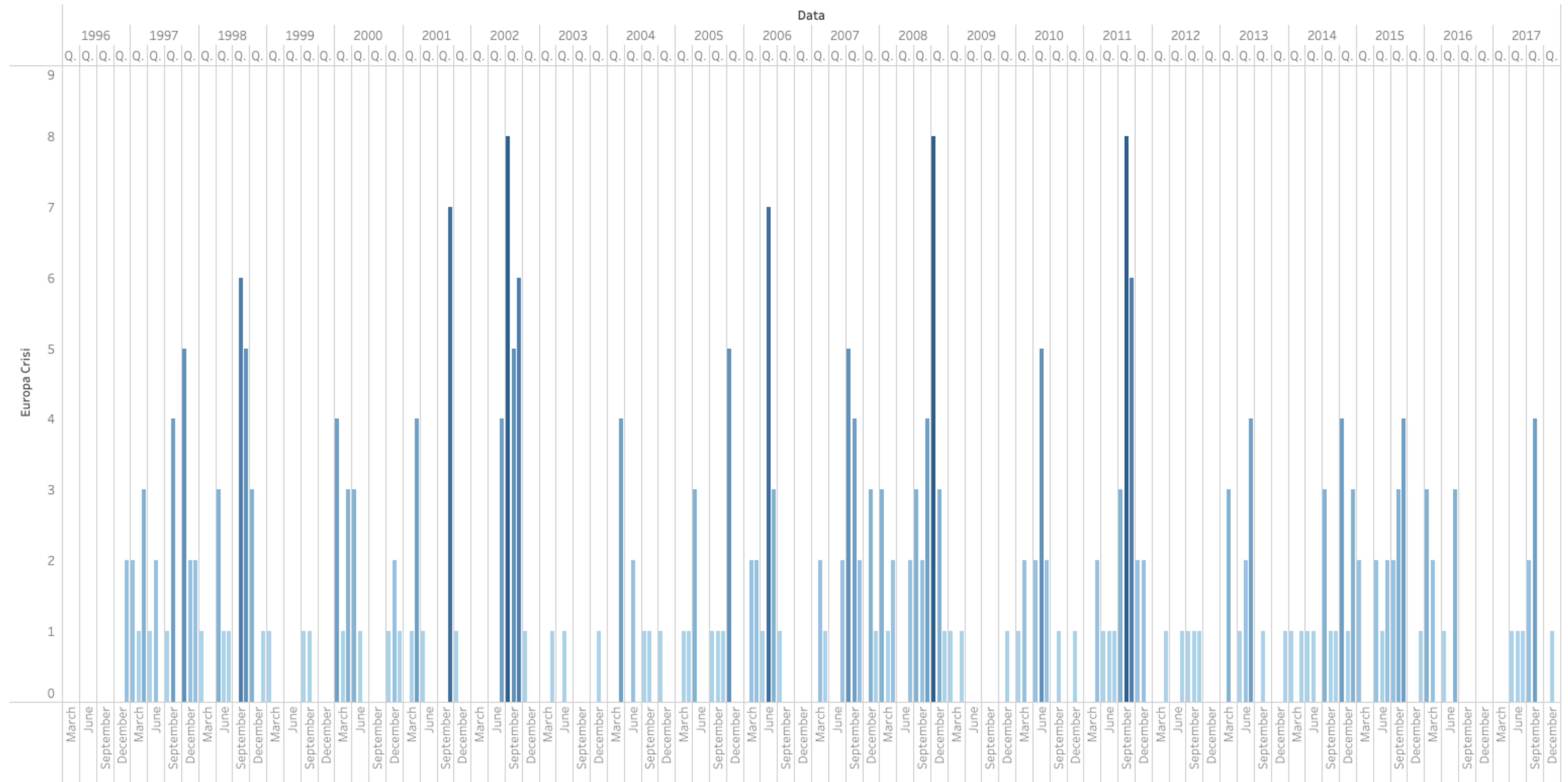
How predict the next financial crisis?

Predicting a financial crisis is certainly not easy.

There are so **many factors** at play. We will see how to select the **most important features** and build a model. And we will see what **the limits** are.

Data

- The Stock Price Index, The yield of the 10-year government bond of 39 countries across the globe
- The exchange rates of 18 currencies against the United States dollar, as well as additional financial indices such as Oil, Gold, and Vix.
- These selected financial indices include the most sensitive market variables, covering the three most important financial markets, namely **America**, **Asia** and **Europe**.
- The data was sourced from the FRED database and the SNL (S&P Global Market Intelligence) website, and covers the period 1996-2017, measured on a daily basis.
- That is, a 22-year period which includes a number of economic crises of different nature and severity



European Financial crisis (over 8 main countries by GDP)

Methodology

- Prepare the dataset
- Selection of the features
- Oversampling
- Prediction

Find the past crisis

- The devaluation of the Mexican Peso (20/12/1994), which triggered the Latin America crisis known as the **Tequila Crisis**
- The devaluation of the Thai Baht (2/7/1997), which triggered the crisis in Asia known as the **Asian Flu**
- The Russian default (17/8/1998), also known as the “Russian Virus,” which caused severe liquidity problems to global markets (Baig & Goldfajn, 2000).
- The beginning of Long Term Capital Management (LTCM) recapitalization (23/9/1998)
- The Hong-Kong stock market crash (28/10/1998).
- The currency devaluation in Brazil (13/1/1999).
- The collapse of **Argentine currency board** (20/12/2001).
- The US and EU **dot.com collapse** (April 2000)
- The Brazilian elections (October 2002)
- The Brazilian run-up to presidential elections (2003)
- **The global financial crisis, which started in stock markets in 2007, and led to a recession in the real sector of the economy (2007).**
- **Greek Sovereign crisis (2010).**
- **European Sovereign crisis (2011, 2012)**
- **US Elections (2016)**
- **Brexit Vote (2016)**

Prepare the dataset

Stock Index	Identify the events crisis	For each day, calculate The lag variable starting from 1 to 60 days
Currency and Bond Markets		
Libor rate	Calculate the returns for each day (log returns)	
Oil price		
Variables	Transformation	LAG

Define a financial crisis

We calculated the initial empirical distribution of returns (pertaining to the first day in our analysis) based on the stock index returns of the first 200 observations (covering the period 10/01/1996–15/10/1996). For each subsequent record (day in the examined period), we recalculated the empirical distribution of returns in order to incorporate the new observation.

An event is identified as "outlier", if the return was below the threshold (first percentile or five percentuile), or of the new empirical distribution.

Define a financial crisis

Let's define a regional and global **crisis events** as the reflects of the negative **co-exceedance** in the stock markets on a given day across countries (for the regional variables) or regions (for the global variable)

We need some thresholds:

- America: At least 3 events per day (out of a total of 7 countries).
- Asia: At least 6 events per day (out of a total of 13 countries).
- Europe: At least 8 events per day (out of a total of 19 countries).
- Global: At least 2 regions are in negative co-exceedance mode on a daily basis.

Feature selections

The constructed dataset comprises an excessive number of independent variables, which is clearly disproportional to the size of the dataset.

The number of days available in the sample; specifically, we are dealing with more 400 variables over around 5400 days (data points).

Fitting a machine learning model to such a huge number of independent variables will suffer of dimensionality problem.

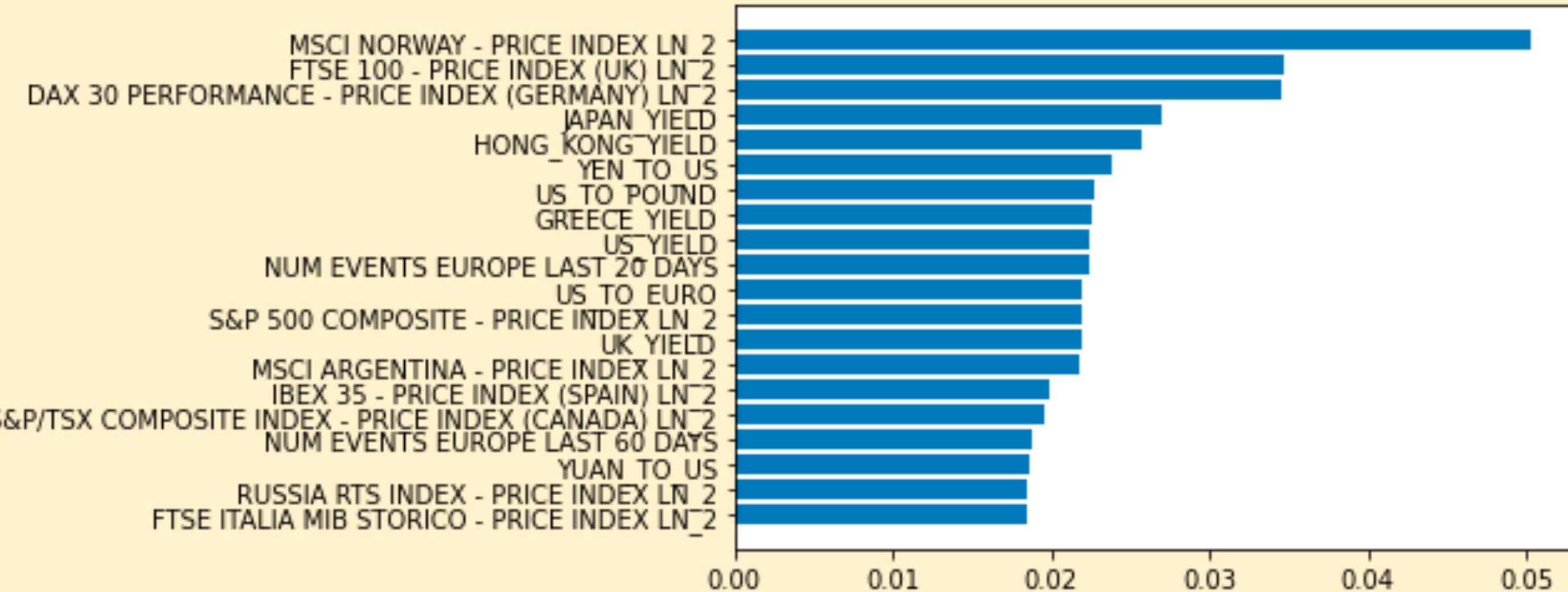
The **Boruta algorithm** is based on a **Random Forest** model. Based on the inferences of this Random Forest, features are removed from the training set, and model training is performed anew.

Boruta infers the importance of each independent variable (feature) in the obtained predictive outcomes by creating shadow features.

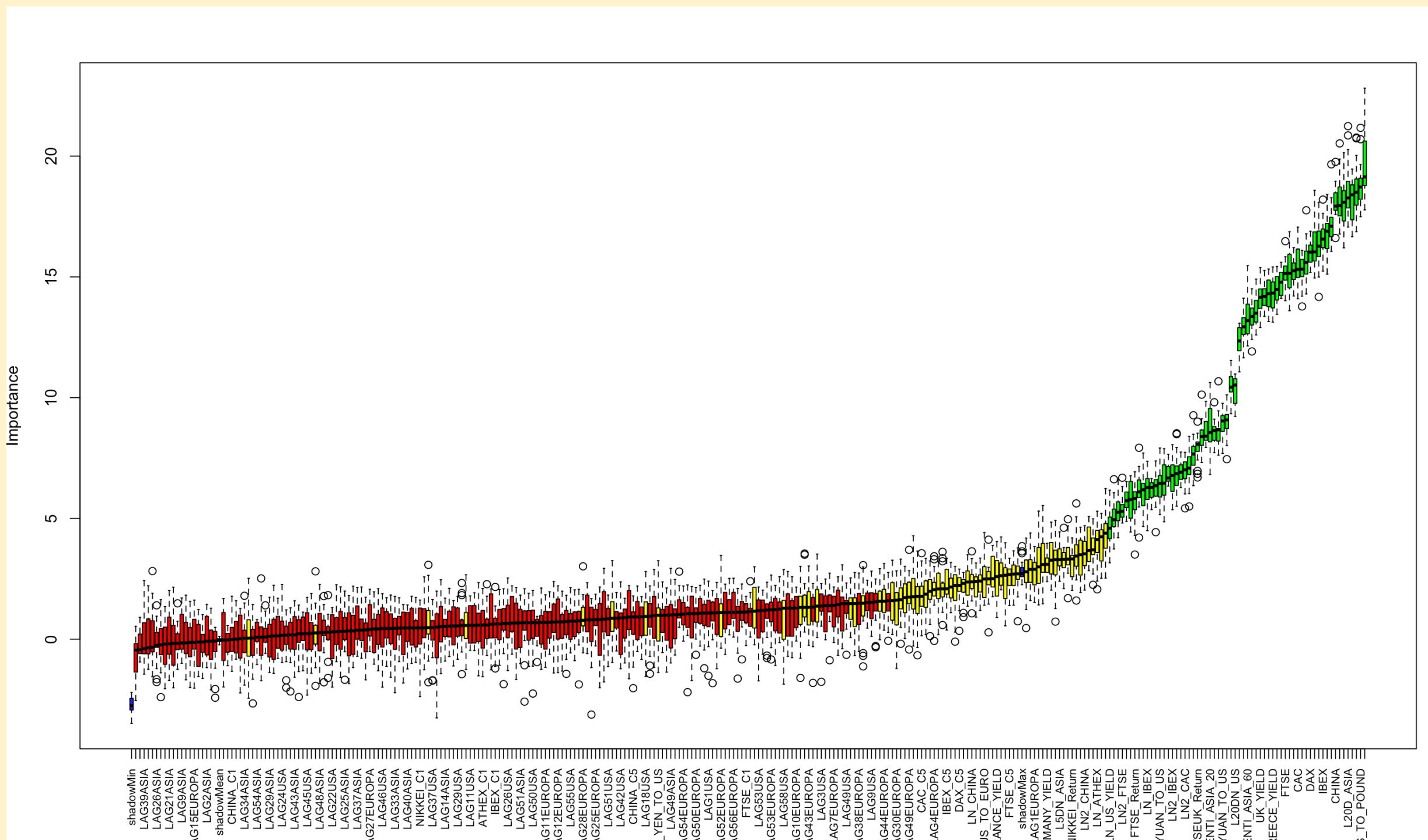
This helps in the identification of all attributes which are in some circumstances relevant for classification, the so-called all-relevant problem.

Random Forests / Boruta

Random Forests



Boruta



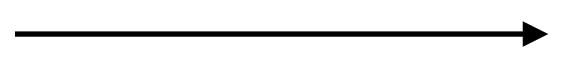
Due to the rather **scarce** nature of global stock crash events, a **very small fraction** of the derived binary dependent variables takes values equal to one in our dataset.

Only around 1% of the observations corresponds to crash events.

With **SMOTE** we will create a balanced artificial datasets, by leveraging sampling methods and a smoothed bootstrap approach.

Oversampling

DATASET



XGBoost

CART DECISION TREE

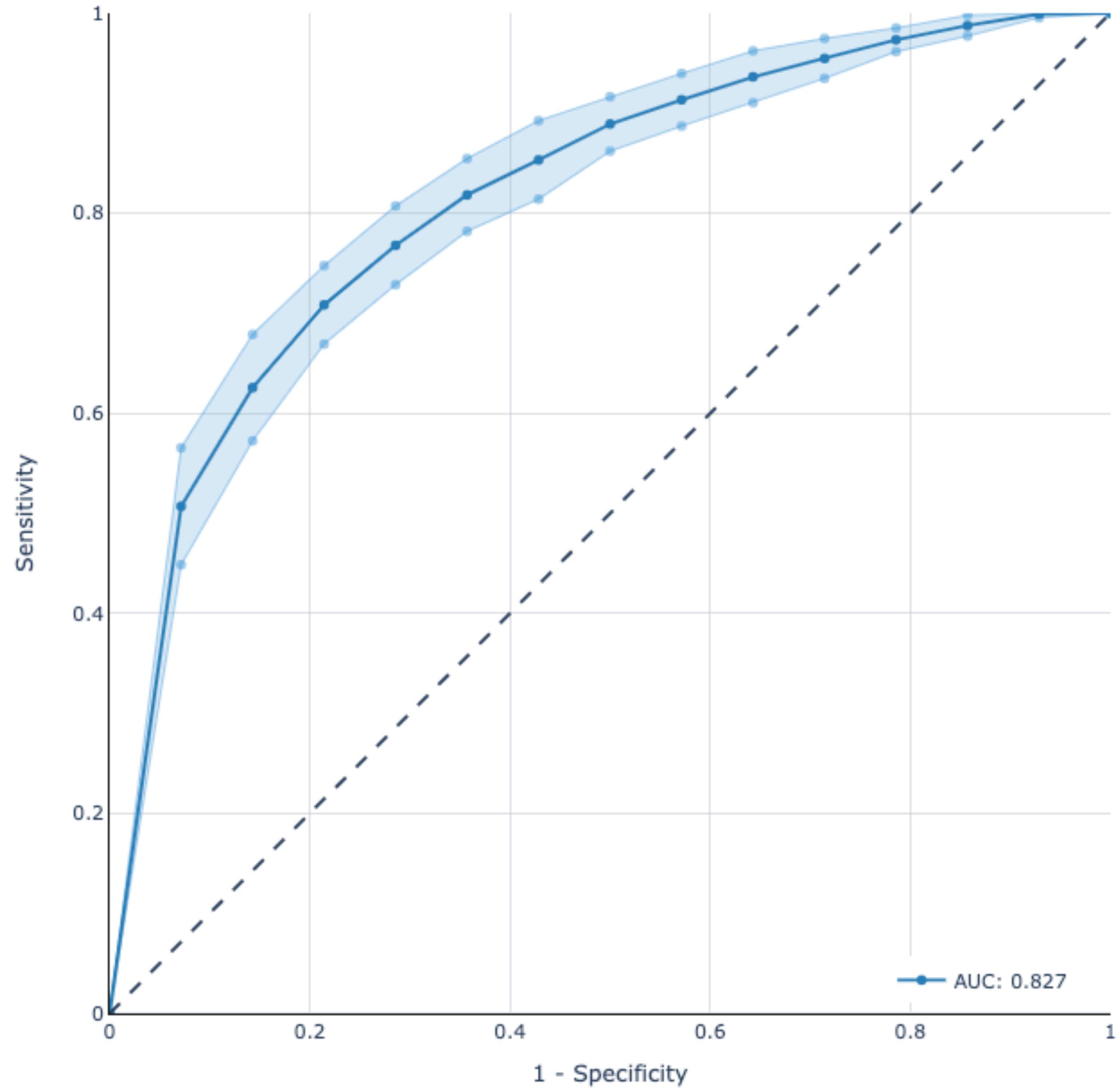
LOGISTIC REGRESSION

<https://github.com/mauropelucchi/pyconit2022>

Are you ready?



Results



Am I here to tell
You that is possible to predict a crisis
only on the next 20 days?

And then?

THE BIG SHORT

They looked
through the nineteen thirties
The housing market collapsed...
Nationwide.

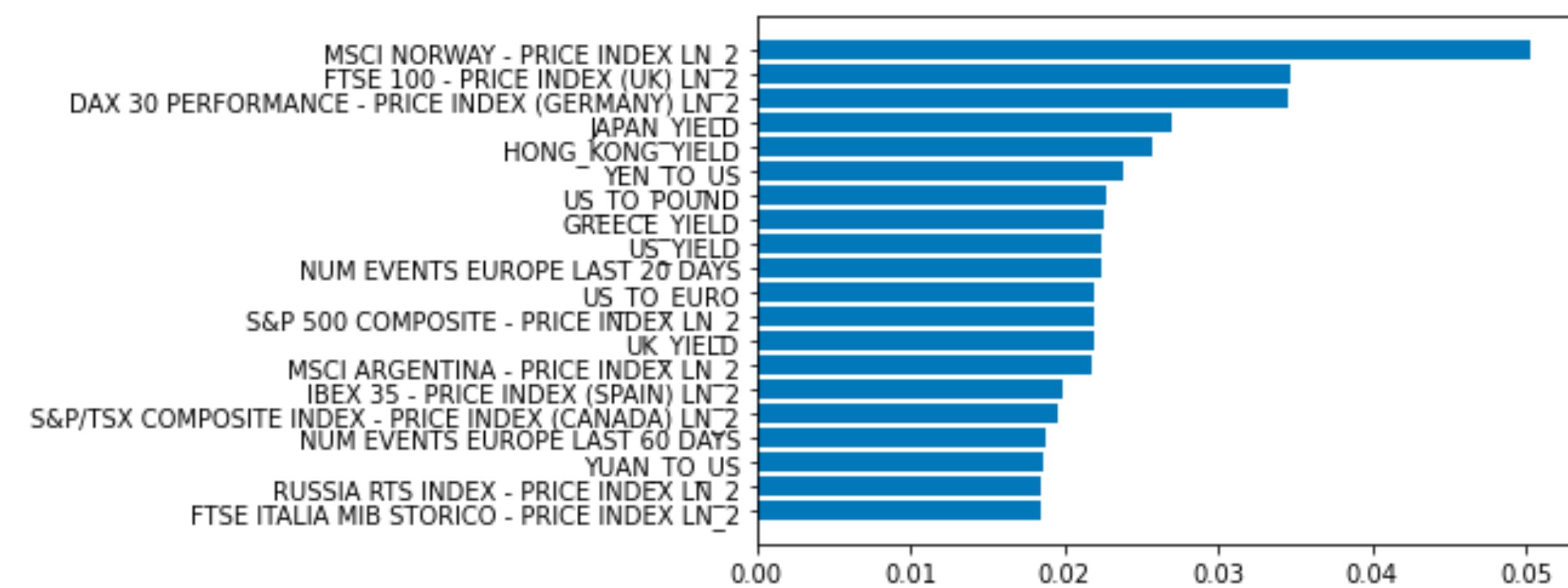
By approximately 8 percent.
It wasn't default.

There were... Very... **Specific identifiers.**

Extremely recognizable.

Wait... for instance
Is that rapid rise and complexity,
and the rates... they are fraud-ed
did you know...
They're going up!

Michael Burry



More than the model, we are interested in the indicators that lead to the prediction