ECB-RESTRICTED

EUROPEAN CENTRAL BANK

BANKING SUPERVISION

Medusa

Increasing consistency in the supervision of banks’ internal models using ML

The views expressed are those of the author and do not necessarily reflect the position of the ECB.

15 March 2023

Hans Gmasz

Agenda

1. BUSINESS BACKGROUND
2. PRODUCT IDEA
3. IMPLEMENTATION
4. DEEP DIVE ON CLASSIFICATION MODEL

In terms of assets under supervision the ECB is the largest banking supervisor in the world[1]

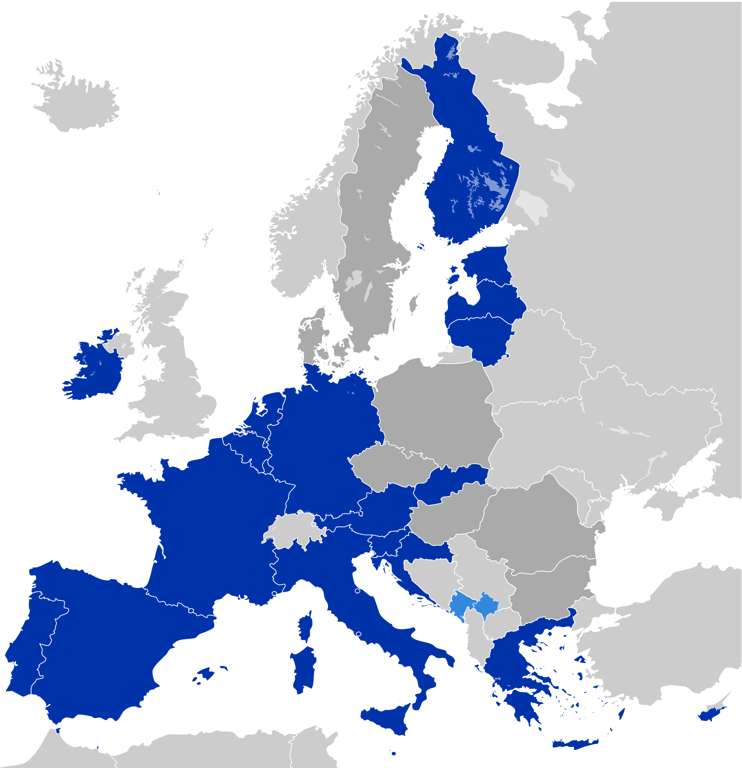
The **Single Supervisory Mechanism (SSM)** refers to the system of banking supervision in Europe. It comprises the ECB and the national supervisory authorities of the participating countries.

All 20 **euro area countries participate automatically** in European banking supervision.

Other EU countries that do not yet have the euro as their currency can choose to participate by entering into **“close cooperation”** with the ECB (Bulgaria since October 2020).

The **ECB directly supervises the 111 significant banks** of the participating countries. These banks hold almost 82% of banking assets in these countries.

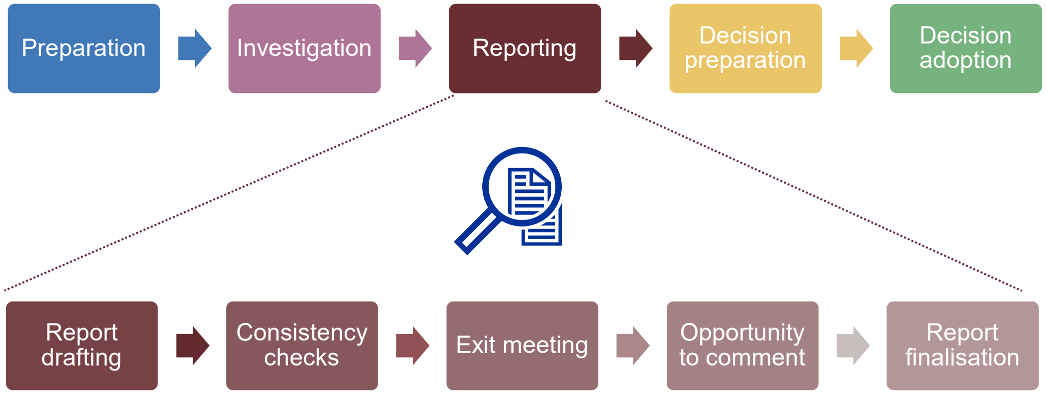
Banks that are not considered significant are known as “less significant” institutions. They continue to be supervised by their national supervisors, in close cooperation with the ECB.[2]



1 https://stats.bis.org/statx/srs/table/b1?m=S&f=pdf
2 https://www.bankingsupervision.europa.eu/about/thessm/html/index.en.html

SSM banks' internal models are mainly supervised through a special kind of on-site inspections called internal model investigations

The ECB grants permissions on the use of internal models by means of **ECB Decisions**, being the outcome of **internal model investigations (IMI)**.



Report drafting and **Consistency checks** ensure reports comply with

- Regulation,
- "Soft law",
- Internal guidance.

While there are governance measures implemented to achieve a uniform interpretation, so far there are almost **no business specific technical tools** available to facilitate report drafting and consistency checks.

https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.osi_guide201809.en.pdf

Drafting and QA of model assessment reports are technically involved and prone to errors

Before

Vetting findings of a model assessment report is cumbersome

In every mission findings need to be agreed with various parties at different stages. This requires a lot of copy/paste and tedious formatting; manual processes prone to errors.

Setting appropriate scope and legal basis is challenging

Reporting findings requires adequate slicing and dicing of shortcomings and selecting the most appropriate legal references. Especially the latter often leads to problems and errors.

Consistency checks on reports are time-consuming

Consistency checks of model assessment reports last 4 weeks or more and require multiple rounds of interaction between the reviewer and the Head of Mission.

Ensuring a level playing field across IMIs is key

Besides internal consistency, model assessment reports must be SSM-wide coherent. This requires comparability of findings over time and across missions.

Medusa aims to offer an efficient way to support the drafting and streamline the QA on model assessment reports

Vetting findings of a model assessment report is cumbersome

Setting appropriate scope and legal basis is challenging

Consistency checks on reports are time-consuming

Ensuring a level playing field across IMIs is key

With MEDUSA

Unified editing and export

Medusa allows to extract findings from documents, edit them in the browser and export them in different formats needed throughout the IMI process.

Reference and granularity checks

The tool checks syntax, existence and adequacy of legal requirements and makes suggestions to merge/split findings according to SSM standards.

Consistency check support

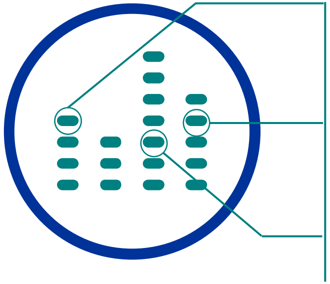
Medusa supports the QA of reports, detecting errors difficult to spot otherwise. This reduces lead time and the number of interaction rounds.

Benchmarking and similarity

Medusa promotes a level playing field by flagging similar findings from the past. It offers advanced semantic search capabilities on a bespoke database of findings.

The Medusa MVP has been released last September, paving the way for the productive tool

Current functionalities



- » **Finding identification and extraction**: Parsing of Model Assessment Reports, detection and extraction of IMI findings
- » **Finding classification**: Multi-labelling of findings on a scheme with currently ~350 different hierarchical labels using advanced machine learning techniques (natural language processing)
- » **Editor and export**: Editing findings directly in the tool and exporting in different Word and Excel formats

- » **Quality assurance checks**: Syntax and existence checks for legal references used (showing full reference texts) and recommendations for merging/splitting of findings
- » **Advanced search**: Browsing a bespoke database of internal model findings using semantic similarity search and advanced filtering

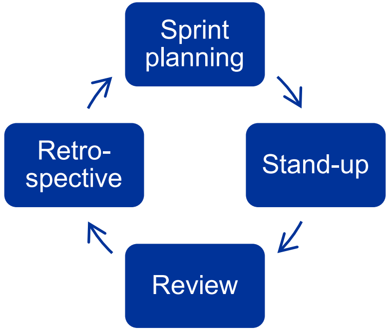
Joint innovation team

- several ECB BAs
- NBB
- BdE

Productive release

Q2 2023

Medusa is being developed in an agile approach and shall be deployed in production in the course of the year



» **Agile development loosely based on scrum**:

- product owner, scrum master and developer roles
- outstanding tasks collected in the product backlog
- work organised in bi-weekly sprints with their individual goals
- stand-up 2x / week; planning, review, retrospective 1x / sprint
- embedded in ECB's SupTech programme

» **Python as the main programming language for the project**:

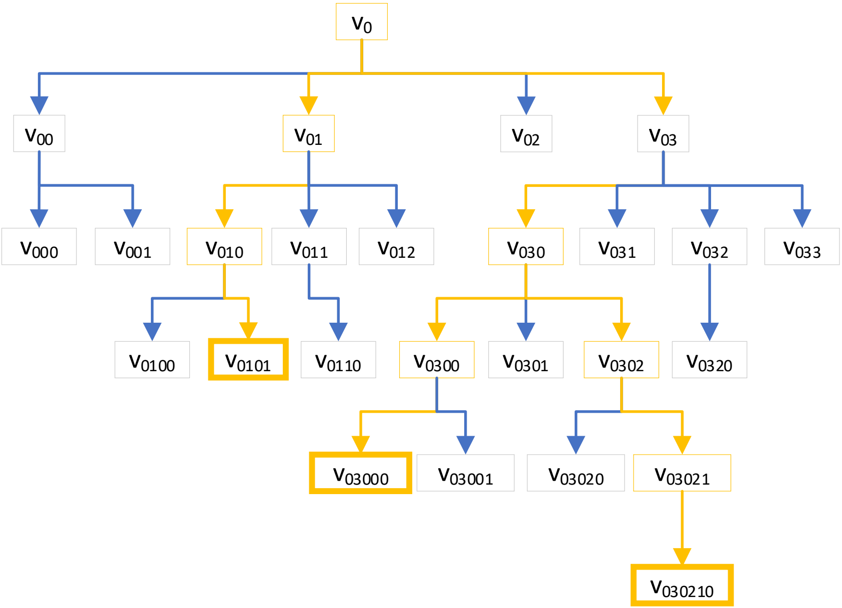
- business logic module (treelib)
- text extraction and processing module (re, docx, bs4)
- ML module (tensorflow, transformers, pandas, numpy)
- API module (connectors, jsonschema, FastAPI)
- APIs implemented using Docker and Kubernetes

» **Product deployment combining multiple services**

- ML components hosted in Azure,
- other backend components hosted in aws,
- UI hosted on a Mendix server

The classification task has to deal with a large label set and only few fully labelled examples to train on

- **Hierarchical multi-class and multi-label learning task**: Data shall be classified according to a labelling scheme that has the structure of a tree with (currently) 350 nodes. Multiple labels can be assigned to one record.
- **A “correct” labelling involves only leaves**: Strictly speaking, an accurate labelling should use only the most granular level of labels available to be as precise as possible.
- **Problem**: Out of the ~10,000 data points available, only ~1,000 are labelled correctly according to this strict rule. Many more do have a partial and cruder labelling.
- **Consider the *closure* of true and predicted labels**: With every true label, all that label’s parent nodes should be considered “almost true” to some degree.



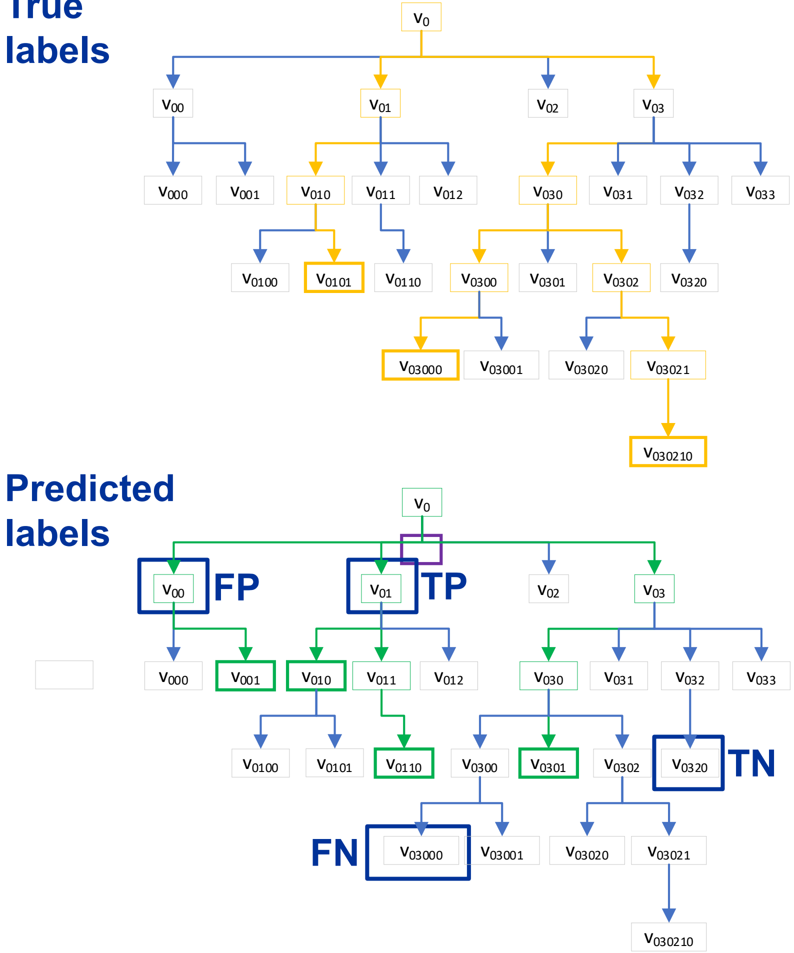
The NN uses a pre-trained BERT, but is highly customised

- **Simple definition of confusion statistics**: After comparing with various more sophisticated confusion statistics based on the Jaccard index (called IoU in image recognition) we decided on the simplest possible definition based on the closure of labels.

- **Loss function and metrics**: Having a confusion matrix at hand allowed us to define a bespoke loss function (based on cross entropy) and performance metrics (we used F1 score as optimising metric).

- **Setup of the NN**: We used a pre-trained BERT model from huggingface to which we attached a fully connected output layer with sigmoid activation; custom loss and metrics were implemented in tensorflow along with some other custom features (tokenizer, regularisation).

- **Training**: We trained with decaying learning rate, first on the fully labelled data, then on all partially labelled data (with very low LR).



Thank you for your attention

» **For questions and comments mail to**: Hans.Gmasz@ecb.europa.eu