**Extraction Approach Description**

# The description provided below will be used to evaluate the approach developed by your team to automatically extract financial data of MNE Groups. This description will be evaluated by the Evaluation panel based on the criteria described in the Evaluation tab of the Extraction Challenge and used for the ranking of your team for the Reusability and Innovativeness Awards.

# **Methodology *(Algorithm reusability and scalability, Data-driven approaches, Availability, quality and maintainability of documentation)***

Please provide a detailed description of the methodology used to develop algorithm-based approaches which automatically extract financial data of MNE Groups. The description should contain (1) the data processing steps, (2) the methods and models used, (3) references to the scientific papers/sources that present the methods and models used, and (4) the time it took to process the data set and extract the relevant financial data.

Bear in mind that the workflow will be also evaluated based on the criteria for the Reusability and Innovativeness Awards.

*This section will be evaluated for:*

*(1) the Algorithm reusability and scalability criterion: likeliness that the described approach can successfully reproduce the solution submitted by the team for the Accuracy award.*

*(2) the Data-driven approaches criterion: evaluated based on whether it is data-driven rather than heuristic. More data-driven approaches will receive higher scores.*

*(3) the Availability, quality and maintainability of documentation criterion: assessed based on the maintainability of the code.*

|  |
| --- |
|  |

## **Architecture *(Architecture)***

Please provide a description of the architecture of your approach. A diagram of the architecture is considered of additional value. Indicate what modifications would be required to apply the approach to similar datasets on a larger scale.

*This section will be evaluated for:*

1. *the Architecture criterion: evaluated based on its modules, their cohesion and their configurability; an architecture which is modular and includes clear connections between modules or components receives a higher score.*

|  |
| --- |
|  |

# **Hardware Specifications *(Algorithm reusability and scalability)***

Please describe the hardware specifications of the machines that were used to run the methodology.

*This section will be evaluated for:*

*(1) the Algorithm reusability and scalability criterion*

**Machine 1**

|  |  |
| --- | --- |
| CPUs | CPU name and capacity |
| GPUs | GPU name and capacity |
| TPUs | TPU name and capacity |
| Disk space | The space required to calculate and store the data |

**Machine 2**

|  |  |
| --- | --- |
| CPUs | CPU name and capacity |
| GPUs | GPU name and capacity |
| TPUs | TPU name and capacity |
| Disk space | The space required to calculate and store the data |

# **Libraries *(Availability, quality and maintainability of documentation)***

Please provide the libraries used for approach, if any, as well as the links to these libraries, if available.

*This section will be evaluated for:*

*(1) the Availability, quality and maintainability of documentation criterion: The use of libraries which are regularly maintained will yield higher scores. (Examples include pytorch, tensorflow, scikit-learn, pandas, numpy, tidyverse, etc.).*

|  |
| --- |
|  |

## **Similarities/differences to State-of-the-Art techniques *(Originality of the approach)***

Please provide a list of similarities and differences between the used methodology and to the state-of-the-art techniques.

*This section will be evaluated for:*

*(1) the Originality of the approach criterion: compare the approach used to the state-of-the-art, i.e. currently published approaches that are closest to the approach applied for the submission, and the extent to which the submission represents an improvement over these approaches.*

|  |
| --- |
|  |

## **Contribution to scientific field *(Future orientation)***

Please describe how your submission contributed to the scientific field, what impact it could have and what could potentially be future work to improve the solution.

*This section will be evaluated for:*

*(1) the Future orientation and impact criterion: the potential effect of the approach used will be evaluated; this includes the scale of impact it has on the problem of extracting financial information from the Internet; the impact will be evaluated based on potential efficiency improvements and cost reductions.*

|  |
| --- |
|  |

## **Lessons Learned *(Future orientation)***

Please state any lessons learned during the competition.

*This section will be evaluated for:*

*(1) the Future orientation and impact criterion: what were the lessons learnt during the competition, and what could potentially be future work to improve the solution.*

|  |
| --- |
|  |

# **Short description of the Team – area of expertise**

Please provide a description of the team, your area of expertise and contact information.

|  |
| --- |
|  |