**Reproducibility Award Methodology Description**

# **Methodology**

Please provide a detailed description of the methodology used for identifying different types of duplicates in the online job advertisement data set. 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 identify the duplicates.

Bear in mind that the workflow will be also evaluated by its originality, interpretability, and simplicity of the methodology.

|  |
| --- |
|  |

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

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

|  |
| --- |
|  |

## **Lessons Learned (optional)**

Please state any lessons learned during the competition.

|  |
| --- |
|  |

# **Hardware Specifications**

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

**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 |

# **Short description of the Team – area of expertise (optional)**

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

|  |
| --- |
|  |