# *PREDICTION OF COLOMBIAN STUDENTS' ACADEMIC SUCCESS USING DECISION TREES ALGORITHMS*

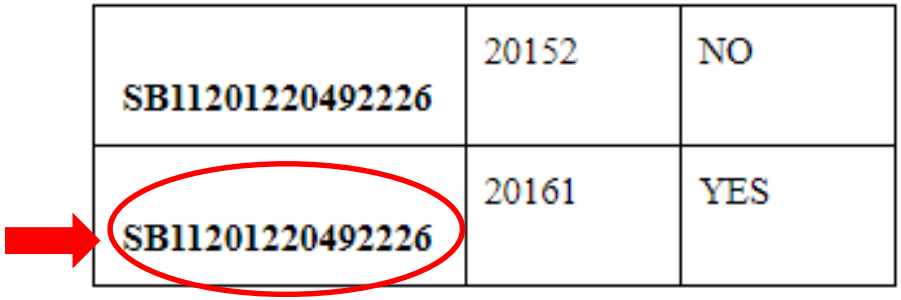
***Juliana Restrepo Tobar***
***David Vergara Patiño***
***Andrea Carvajal Maldonado***
*Medellín, fecha*

# *Designed Data Structure*

***Figure 1:*** DataFrame of students and their information. The student's code is the key and all information such as ID, age, genre is they key value.

| | Key | Key Value | | |
|---|---|---|---|---|
| 0 | **Student´s code** | Period | Studied abroad? | … |
| 1 | **SB11201220492225** | 20152 | YES | … |
| 2 | **SB11201220492224** | 20131 | NO | … |
| 3 | **SB11201220492226** | 20151 | NO | … |

# *Data Structure Operations*



| | | |
|---|---|---|
| **SB11201220492226** | 20152 | NO |
| **SB11201220492226** | 20161 | YES |

**Figure 2:** Information access operation

| | **Complexity** | |
|---|---|---|
| **Method** | **Best case** | **Worst case** |
| iloc/loc | O(1) | O(1) |
| pandas.concat | O(1) | O(1) |

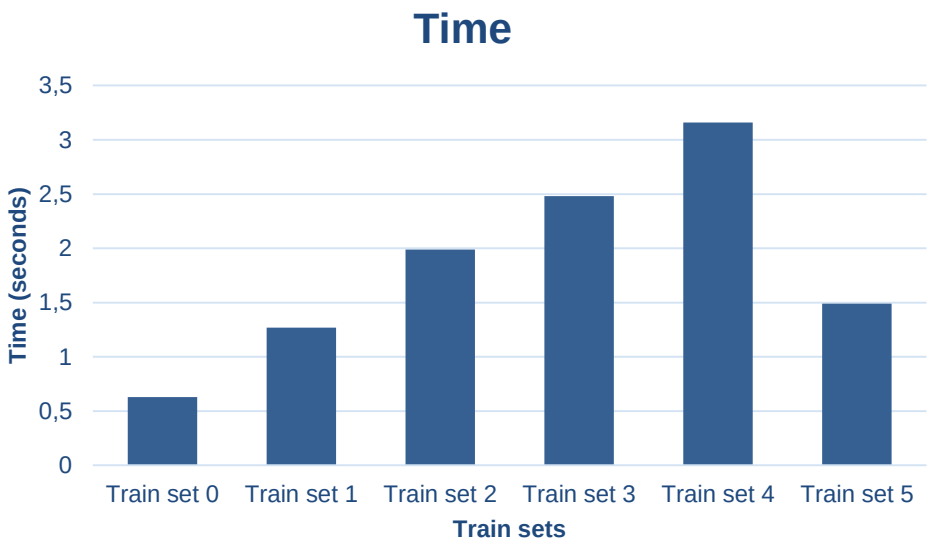
**Table 1:** Complexity of some operations in DataFrame

# *Design Criteria of the Data Structure*

- Since the format of the input data is a matrix, it is very convenient to store it in a table-like structure
- This data structure allows storing different data types
- The DataFrame is one of the most used data structures when working with big data and AI
- It is an efficient and easy way of organizing and managing a large volume of information

# *Time and Memory Consumption*

**Figure 3:**Time taken to create the data



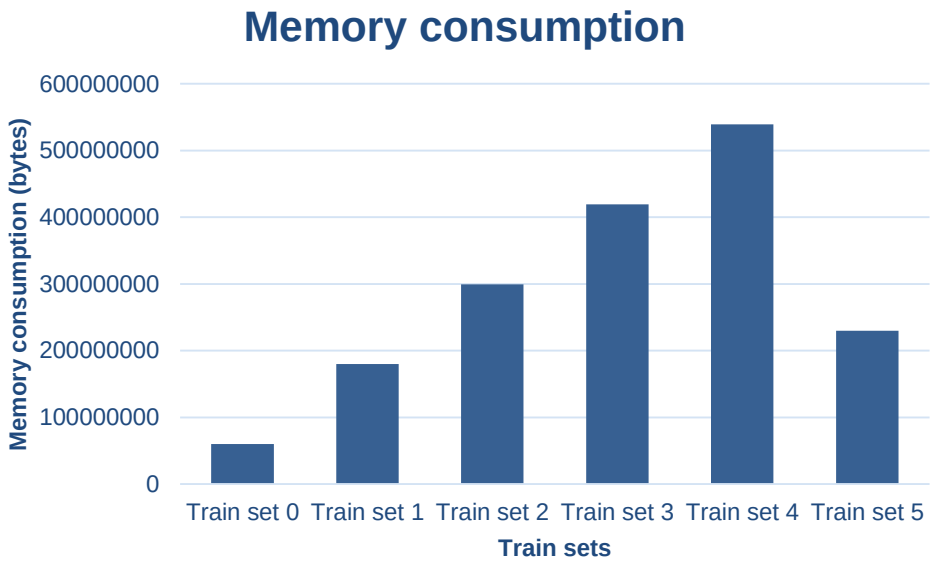




**Figure 4:** Memory consumption of the data

# *Implementation*

# *Report in arXiv*