Airline Departure Data Analysis and Regression

Lucchi Manuele & Tricella Davide

August 12, 2022

Abstract

The purpose of this paper is to evaluate the usage of a Logistic Regression model on a airlines dataset to predict flight cancellation or diversion, in a scalable and time/space efficient implementation.

1 Definitions

Stoichiometry The relationship between the relative quantities of substances taking part in a reaction or forming a compound, typically a ratio of whole integers.

Atomic mass The mass of an atom of a chemical element expressed in atomic mass units. It is approximately equivalent to the number of protons and neutrons in the atom (the mass number) or to the average number allowing for the relative abundances of different isotopes.

2 Dataset

- Overview dataset - Struttura delle colonne - Divisione dei file - Numero di record - Quali colonne prendiamo - Quanti record prendiamo

3 Preprocessing Techniques

- balancing - normalization - splitting - dates and enums to numbers

4 Preprocessing Parallelization

5 Model

6 Performances

7 Experiments

8 Results and Conclusions

The atomic weight of magnesium is concluded to be $24\,\text{g}\,\text{mol}^{-1}$, as determined by the stoichiometry of its chemical combination with oxygen. This result is in agreement with the accepted value.

Figure 1: Figure caption.