

# INTRODUCTION TO MACHINE LEARNING AND DATA MINING

---

**Student names and ids:** Anders H. Opstrup (s160148); Gu Jinshan (s161944)

**Report:** 1

---

## Description of the data set

- Describe if the attributes are discrete/continuous, Nominal/Ordinal/Interval/Ratio.

1. X - Discrete, Nominal
2. Y - Discrete, Nominal
3. month - Discrete, Ordinal
4. day - Discrete, Ordinal
5. FFMC - Continuous, Interval
6. DMC - Continuous, Interval
7. DC - Continuous, Interval
8. ISI - Continuous, Interval
9. temp - Continuous, Interval
10. RH - Continuous, Ratio
11. wind - Continuous, Ratio
12. rain - Continuous, Ratio
13. area - Continuous, Ratio

- Give an account of whether there are data issues (i.e. missing values or corrupted data) and describe them if so.

There is no missing values or corrupted data.

- Describe the basic summary statistics of the attributes.

Statistics	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
Mean	90.64468	110.8723	547.94	9.021663	18.88917	44.2882	4.017602	0.021663	12.84729
Median	91.6	108.3	664.2	8.4	19.3	42	4	0	0.52
Variance	30.41268	4094.018	61417.81	20.74862	33.65168	265.7448	3.20381	0.087422	4044.226
Standard deviation	5.51477	63.98451	247.8262	4.555065	5.801007	16.30168	1.789919	0.295673	63.59423