

Model construction methodologies

Yacine Debbabi

December 1, 2020

Contents

1	Exploratory data analysis	1
----------	----------------------------------	----------

1 Exploratory data analysis

Is the dataset clean?

- Identify missing values using `df.isna().sum()`.
- Plot feature distributions to identify complications (e.g., outliers, clippings, rounding errors, non-normal distributions).

Which features bring information?

- Plot the features against the response.
- Identify which features are correlated between themselves.