

Examen de Machine Learning & Text Mining

Etude de cas en management de projet :

Dans cette étude, vous allez concevoir et développer un modèle intelligent pour prédire l'effort de développement d'un projet logiciel. Le dataset ci-dessous contient les données concernant les projets de développement logiciel déjà achevés.

Story Point : la taille du projet en story point

Velocity : la quantité (en story point) que l'équipe peut développer pendant une période.

Team Size : l'effectif de l'équipe de développement

Story Point	Velocity	Team Size	Effort
156	2.7	5	63
202	2.5	5	92
173	3.3	5	56
331	3.8	5	86
124	4.2	5	32
339	3.6	7	91
97	3.4	5	35
257	3	5	93
84	2.4	5	36
211	3.2	5	62
131	3.2	5	45
112	2.9	5	37
101	2.9	5	32
74	2.9	7	30
62	2.9	5	21
289	2.8	5	112
113	2.8	5	39
141	2.8	5	52
213	2.8	5	80
137	2.7	5	56
91	2.7	5	35

Travail à faire :

1. Chercher le meilleur classifieur en menant une étude empirique comparative entre les techniques de machine learning (travaillées dans les séances de TP).
2. Les livrables sont : rapport + code source Python

Dans le rapport,

- il faut suivre une démarche adéquate avec un projet machine learning : collecte données, analyse et description du dataset...
- N'hésiter pas à utiliser toutes les techniques permettant d'améliorer la performance du modèle (feature selection, optimisation,...)
- L'évaluation est basée sur deux indicateurs (en italique)

$$AbE_i = |y_i - \hat{y}_i|$$

$$MRE = \frac{AE_i}{y_i}$$

$$Pred(25\%) = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 & \text{if } MRE_i \leq 25\% \\ 0 & \text{otherwise} \end{cases}$$

$$MMRE = \frac{1}{N} \sum_{i=0}^N MRE$$

Où y_i est l'effort réel, \hat{y}_i représente l'effort estimé du projet numéro i

- La base de test est composée de 30% du dataset.
- La comparaison doit inclure aux moins 3 techniques vues en cours. Toutefois, vous pouvez utiliser d'autres techniques.