Evaluation

Daten Vorbereitung

Pour chaque méthode, les données seront préparées de différentes manières. Pour les méthodes se basant sur les réseaux neurones, les données sont préparées avec les scalers suivant : pas de scaler, MinMaxScaler, StandardScaler, BooleanScaler.

Pour les méthodes se basant sur les clusters, les données sont préparées avec plus de scalers car il est plus facile et rapide de faire la comparaison avec ces méthodes qui ont l’avantage de ne pas passer par des phases d’apprentissage chronophage.

Pas de scaler signifie qu’aucune préparation n’est faite sur les données et sont donc données directement aux méthodes.

MinMaxScaler: Skaliert den Datensatz neu, sodass alle Feature-Werte im Bereich [0, 1] liegen.

Standard Scaler: Entfernt den Mittelwert und skaliert die Daten auf Einheitsvarianz.

MaxAbsScaler: Skaliert so, dass die Trainingsdaten im Bereich [-1,1] liegen, indem jeder Datenpunkt durch den Maximalwert geteilt wird.

RobustScaler: Die Zentrierungs- und Skalierungsstatistik vom RobustScaler basiert auf Perzentilen und wird daher nicht von einigen wenigen sehr großen Randausreißern beeinflusst.

PowerTransformer: Wendet eine Leistungstransformation auf jedes Feature an, um die Daten Gauß-ähnlicher zu machen, um die Varianz zu stabilisieren und die Schiefe zu minimieren. Derzeit werden die Yeo-Johnson- und Box-Cox-Transformationen unterstützt und der optimale Skalierungsfaktor wird bei beiden Methoden über die Maximum-Likelihood-Schätzung bestimmt.

QuantileTransformer: Wendet eine nichtlineare Transformation an, sodass die Wahrscheinlichkeitsdichtefunktion jedes Merkmals auf eine gleichmäßige oder Gaußsche Verteilung abgebildet wird.

Normalizer: Skaliert den Vektor für jede Probe neu, um unabhängig von der Verteilung der Proben eine Einheitsnorm zu erhalten.

Training

Ergebnis