**Skip-gram** est l'une des techniques d'apprentissage non supervisé utilisées pour trouver les mots les plus liés pour un mot donné.

**Algorithmes Word2vec**  
Alors comment ça a été fait? La représentation word2vec est créée à l'aide de 2 algorithmes: le modèle Continuous Bag-of-Words ( **CBOW** ) et le modèle **Skip-Gram** .

**StratifiedShuffleSplit**

   Validateur croisé ShuffleSplit stratifié  
Fournit des indices de train / test pour diviser les données dans des ensembles de train / test.  
Cet objet de validation croisée est une fusion de StratifiedKFold et ShuffleSplit, qui renvoie des plis aléatoires stratifiés. Les plis sont réalisés en conservant le pourcentage d'échantillons pour chaque classe.

**GridSearchCV**

   Recherche exhaustive sur les valeurs de paramètres spécifiées pour un estimateur.  
Les membres importants sont en forme, prévoyez.  
GridSearchCV implémente une méthode «fit» et une méthode «score». Il implémente également «prédire», «prédire\_proba», «fonction\_décision», «transformer» et «inverser\_transform» s'ils sont implémentés dans l'estimateur utilisé.

**train\_test\_split**

Diviser les tableaux ou matrices en sous-ensembles aléatoires de train et de test

**word2vec**  
**word2vec** est un concept bien connu, utilisé pour générer des vecteurs de représentation à partir de mots.  
Il existe de nombreux bons tutoriels en ligne sur word2vec, comme celui- ci et celui- ci, mais décrire **doc2vec** sans **word2vec** manquera l’essentiel.