L’algorithme de classification bayésienne est un outil de classification qui s’exprime par : *p(m|c)=n(m,c)/n(c))* avec, dans notre cas, c une classe (positif, negatif ou neutre), m un mot, n(c) le nombre total de tweets de la classe c, n(m, c) le nombre tweets qui contient mot m dans la classe c et P(m|c) et le nombre d’occurrence du mot m dans un texte de classe c et p(c) est .le radio de tweets neutres dans tous les tweets de la base de tweets.

Pour connaitre la classe c d’un tweet, on le divise en mot dans un premier temps, dans un second temps on cherche le nombre d’occurrence des tweets qui contiennent ce mot m. Puis on multiple les occurrences de tous les mots par la proportionnalité de tweets de classe c de la base de tweets. Si le mot n’apparait jamais dans la classe c, on remplace le résultat nul par (p(m|n)+1)/n(c)+N. Ensuite on utilise le résultat pour multiple avec p(c), on peut obtenir le probabilité de ce tweet qui apparait dans la classes neutre. Comparez avec les probabilités de trois classes, on peut annoter ce tweet par la classes qui a la probabilites le plus grande.