

در قسمت Naïve Bayes اثر Bag Of Words دیده می شود چون دیتای ما مربوط به دو کلاس انسان و غیر انسان است که به دلیل اسامی انسان ها یا اسامی غیر انسان ها همچنین با توجه به لغات و اصطلاحات مشترکی که انسان ها یا غیر انسان ها به کار می برند که هر کدام از این دسته ها در بین اعضای همان دسته Bag Of Words دارند.

در نظر گرفتن ترتیب کلمات تاثیر به سزای نداشته ولی در تفاوت بین جملات کوتاهتر و بلندتر گاهی باعث اشتباه شده است. گاهی پیش می آید که مثلا انسان ها اسم غیر انسان ها را به کار ببرند که این باعث دسته بندی اشتباه می شود. در مواردی که احتمال هر دو کلاس نزدیک به هم است عوامل گفته شده باعث دسته بندی اشتباه یا درست می شود.

مثال:

Human fate aint written artanis

Human -13.821 NotHuman -13.616

در این مثال می بینیم به خاطر کلمه artanis که یک اسم غیر انسان است و در کلاس NotHuman متداول است باعث دسته بندی اشتباه شده است.

در قسمت MaxEnt در مقایسه با قسمت Naïve Bayes تاثیر روی Accuracy نداشته و نزدیک به همان دقت قبلیست چون تقریبا همان احتمالات کلمات را به عنوان ویژگی گرفته ایم و چون دقت بالای نیم و حدود ۰/۷ بود دیگر نیاز به اضافه یا حذف کردن ویژگی دیده نشد.

Naïve Bayes : Accuracy = 0.713

MaxEnt : accuracy = 0.7033333333333334

استفاده از الگوریتم پیچیده تر زمانی بهینه تر است که ویژگی هایمان را طوری بگیریم که Overfit نداشته باشند.