

Bu yazının amacı, bilgisayar bilimlerinde özellikle istatistiksel doğal dil işleme (Statistical Natural Language Processing, NLP) ve metin madenciliği (text mining) , bilgi getirimi (information retrieval) gibi konularda geçen vektör uzayı modelini (vector space model) açıklamaktır. Bu modelleme yöntemine göre herhangi bir metin veya metin parçasını vektörel olarak göstermek ve bir uzay içerisinde modellemek mümkündür. Örneğin bir yöneyi (vektör, vector) her boyuttaki değeri ile göstermemiz gerekir. Misal olarak [1,2] vektörünü iki boyutlu uzayda aşağıdaki şekilde çizilebilir:

İki boyutlu uzayda ilk boyut ve ikinci boyut yukarıdaki temsili resimde gösterilmiştir. Buna benzer şekilde bir vektörün sonsuz sayıda boyutu olabilir. [1,2,1,3,2,1,2,3,2,1,2,2,1,3] şeklinde 14 boyutlu bir vektörden bahsedebiliriz. Bu vektörün gösterimi veya çizilmesi gerçekten gerekmemektedir çünkü amaç birbiri ile ilişkisi olmayan boyutlarda farklı değerlere sahip bir vektörü temsil etmektir. Buradaki önemli nokta bu boyutların arasında ilişki bulunmamasıdır. Yukarıdaki temsili şekilde de ilk okulda öğrenilen ve birbirine dik olan eksenlerin ilişkilerinin olmaması kabulüne dayanılarak çizim yapılmıştır. Diğer bir deyişle birinci boyutun kaç olduğunun 2. boyuta veya ikinci boyutun kaç olduğunun 3. boyuta bir etkisi yoktur.