1. Machine Learning adalah suatu computer atau program computer yang mengalami peningkatan seiring pembelajaran pada contoh ataupun data set sehingga developer tidak perlu susah dalam coding pemrograman.
2. SAVMAP <http://lasig.epfl.ch/savmap> . Salah satu objective pada SAVMAP adalah classification hewan semi-arid savanna di Namibia dengan image processing yang diambil dengan Unmanned Aircraft Vehicle (UAV). Mereka mencoba beberapa metode machine learning classifier diantaranya LOG , SVM with an RBF Kernel, NB, dan RF dan mencoba mengcomparasinya dengan hasil performasi 80% presisi dan recall menggunakan HOG features untuk feature set.
3. ~

A. 0 – 1 jika input bukan non negative dan (-1 – 1) jika input negative.

1. Ya sama, karena range nilai yang terdapat pada similarity adalah dari 0 – 1 dimana 0 menyatakan kedua data tidak mirip dan semakin besar tingkat kemiripan akan menyatakan data tersebut sama.
2. Jika kita kurangkan rata-rata dari x dan y maka pengukuran menjadi invariant dan hasilnya adalah titik tengah (Pearson Product Moment Correlation Coefficient) PPMC dimana sudutnya adalah pusat dari vector.
3. Euclidean :

Jika panjang = 1, maka jaraknya ke 0 adalah

Jika hasilnya 1 maka kuadratnya juga 1:

Kemudian pada rumus Euclid yang diekspansi, selanjutnya subtitusi

Cosine Similarity :

Jika panjang norm = 1 maka

Karena nilai cosine similarity selalu diantara 0 dan 1, maka kita bisa defenisikan jarak cosine dengan :

Misalnya kita punya vector v dan w dengan menggunakan euclid, dan v lebih dekat ke suatu point p daripada w.

Sehingga, kita dapat hubungan antara cosine similarity dan Euclidean

1. Euclidean

Correlation

jika X dan Y distandarkan, mereka masing-masing akan memiliki rata-rata 0 dan standar deviasi 1, sehingga rumus menjadi

Kembali ke rumus Euclidean jika x dan y distandarkan maka keduanya = n. Sisa Σxy sebagai non-konstan, maka formula bisa disederhanakan menjadi :

1. ~
2. Pertama berdasarkan pairwise proximity bisa menggunakan minimum pairwise similiarity atau maximum pairwise dissimilarity. Kedua bedasarkan data di Euclidean kemudian hitung total atau rata-rata jarak data ke centroid.
3. Dengan menghitung jarak antara centroids dari kedua kumpulan data.
4. Dengan menghitung rata-rata kedekatan pasangan dari kumpulan data di satu kelompok group dengan kumpulan data lain. Atau dengan mengambil kedekatan yang maksimal atau minimum.
5. ~
6. Untuk satu document mempunyai bobot maximum yaitu log m.
7. Kata yang muncul pada setiap document memiliki bobot 0
8. Pengamatan dari kata yang ada di setiap document tidak mempengaruhi document lain.
9. Near Duplicate Image Detection: min-Hash and tf-idf Weighting <https://www.robots.ox.ac.uk/~vgg/publications/papers/chum08a.pdf>