

Opinto-ohjelmani on tietojenkäsittelytieteen kandidaatti (TKT). Tulen käyttämään projektin dokumentaatioissa Suomen kieltä. Käytän työssä Pythonia. Pythonin lisäksi hallitsen siinä määrin Fortran:ia että voisin vertaisarvioida sillä tehtyjä projekteja.

Työn aiheena on hahmontunnistus, käsin kirjoitettujen numeroiden tunnistus, käyttäen  $k$ :n lähimmän naapurin menetelmää ( $k$ -NN). Toteutan työssäni luokitteluun käytettävän, koneoppimis algoritmin,  $k$ -NN algoritmin. Algoritmin tekemä luokittelu, algoritmin ydin, perustuu etäisyyksien laskemiseen sekä ennusteen tekemiseen laskettujen etäisyyksien perusteella. Algoritmin ydin perustuu ensimmäisessä versiossa euklidiseen etäisyyteen, kun algoritmi toimii riittävän hyvin tällä etäisyyksimitalla, lisään etäisyyksimittoja joita on esitetty lähteessä (Dubuisson & Jain, 1994).

Työssäni käytän python array-tietorakennetta, joka on taulukkomuotoinen tietorakenne. Käytettävät tietorakenteet tarkentuvat työn edetessä. Käyttöliittymäksi algoritmille olen ajatellut www-sivua jossa numeron tunnistuksen voi tehdä lataamalla kuvan numerosta sivulle.

Toteutettu algoritmi tunnistaa numeroita, eli ratkaisee ongelman: mikä numero on käsin kirjoitettu. Syötteenä algoritmi saa kuvan käsin kirjoitetusta numerosta, sekä koulutukseen käytettävän datan MNIST tietokannasta.

Tavoitteena on toteuttaa algoritmi, jonka aika- ja tilavaativuus on  $O(n)$ . Aikavaativuus  $O(n)$  onnistunee, jos mikäli algoritmin ydin, etäisyyksien laskeminen, perustuu euklidiseen etäisyyteen. Etäisyyden laskennassa tulen käyttämään myös muutamaa toista etäisyyksimittaa, mikä saattaa korottaa aikavaativuutta.

Lähteet, joita on tarkoitus käyttää, lista päivittyy työn edetessä.

Dubuisson, M.-P., & Jain, A. K. (1994). A modified Hausdorff distance for object matching. Proceedings of 12th International Conference on Pattern Recognition, 566–568.

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=576361>

$k$ -nearest neighbors algorithm. (2025). viewed 2 September 2025, [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)

Harjoitustyön testaaminen. (2025). viewed 2 September 2025, <https://algotestlab.github.io/testing-fi>