

Aineiston nimi: Virtavesien (alle 5 m) ennustettu puromaisuus

Dokumentti päivitetty 19.8.2025

Lyhyt kuvaus: Maastotietokannan virtavesiviiva-aineistolle on laskettu ominaisuustietoja ja arvioitu puromaisuuden todennäköisyys kahdella eri koneoppimismallilla. Pääasiallisena tarkoituksena on ollut erottaa kaivetut ojat luonnonuomista. Aineisto kattaa koko Suomen. Aineiston tuottamiseen tarvittava tarkempi metodiikka löytyy Pythonilla kirjoitettuna osoitteesta <https://github.com/PossibleSolutions/StreamClassifier> (Part1)

Aineistoformaatti: GeoPackage

EPSG: 3067, ETRS89 / TM35FIN(E,N)

Aineiston koko: koko Suomen aineisto 8,2 GB

Saatavuus: Aineiston saa pyytämällä erillisellä tietojensiirto-ohjelmalla aineiston koon vuoksi. Poistetaan tämä kohta, jos aineiston saa suoraan jakeluun.

Aineiston oletusulkoasu: perustuu predicted_class_mean -attribuuttiin.

Aineiston attribuuttitiedot:

Muuttuja	Tyyppi	Kuvaus
fid	kokonaisluku, 64bit	Vektorin yksilöivä koodi
length	desimaali	Viivan pituus
length_simplified	desimaali	Yksinkertaistetun viivan pituus (Distance, Douglas-Peucker) kun suoran viivan poikkeaman etäisyysarvo on 10 m
length_ratio	desimaali	Sinuosity-arvo eli mutkaisuusarvo, tapa 1: laakson keskipituuden arvo (viivan pituuden suhde yksinkertaistetun viivan pituuteen, length/length_simplified)
num_vertices	kokonaisluku, 32bit	Solmupisteiden lukumäärä
max_distance	desimaali	Solmupisteiden maksimietäisyys (etäisyys minkä tahansa viivan solmupisteiden välillä). Jos viiva on täysin suora, maksimietäisyys on yhtä kuin viivan pituus (length).
mean_distance	desimaali	Solmupisteiden keskietäisyys
sinuosity	desimaali	Sinuosity-arvo eli mutkaisuusarvo, tapa 2: viivan pituus jaettuna viivan

Muuttuja	Tyyppi	Kuvaus
		alku- ja loppupisteen suoralla (lyhimmällä) etäisyydellä
Neighbor_DistanceSum	desimaali	Viivan keskipisteen ja lähimmän 20 viivan keskipiteiden etäisyyden summa
Neighbor_DistanceMean	desimaali	Viivan keskipisteen ja lähimmän 20 viivan keskipiteiden etäisyyden keskiarvo
DTW1ha	teksti	Depth To Water 1ha (DTW, https://ckan.ymparisto.fi/dataset/dtw-kosteusindeksikartat) indeksin arvo viivan jokaisessa solmukohdassa listana
DTW1ha_mean	desimaali	Depth To Water 1ha (DTW, https://ckan.ymparisto.fi/dataset/dtw-kosteusindeksikartat) indeksin keskiarvo laskettuna viivan jokaisen solmukohdan arvosta
DTW1ha_median	desimaali	Depth To Water 1ha (DTW, https://ckan.ymparisto.fi/dataset/dtw-kosteusindeksikartat) indeksin mediaaniarvo laskettuna viivan jokaisen solmukohdan arvosta
slope	teksti	Kaltevuus viivan jokaisessa solmukohdassa listana
slope_mean	desimaali	Keskikaltevuus laskettuna viivan jokaisen solmukohdan kaltevuusarvosta
slope_median	desimaali	Mediaanikaltevuus laskettuna viivan jokaisen solmukohdan kaltevuusarvosta
soil	teksti	Maalajiluokka (rasteriaineistossa, joka yhdistelmä GTK:n maaperäaineistoja, 1:20 000, 1:50 000 ja 1:200 000) jokaisessa viivan solmukohdassa listana. Sisältää tiedon sekä pohja-, että pintamaalajista.
soil_surface	kokonaisluku, 32bit	Yleisin pintamaalajiluokka (rasteriaineistossa, joka yhdistelmä GTK:n maaperäaineistoja, 1:20 000, 1:50 000 ja 1:200 000) laskettuna viivan solmukohtien arvojoukosta

Muuttuja	Tyyppi	Kuvaus
soil_base	kokonaisluku, 32bit	Yleisin pohjamaalajiluokka (rasteriaineistossa, joka yhdistelmä GTK:n maaperäaineistoja, 1:20 000, 1:50 000 ja 1:200 000) laskettuna viivan solmukohtien arvojoukosta
ref_class	kokonaisluku, 32bit	Referenssiaineiston luokitus viivalle, 0=oja, 1=puro, nodata= viiva ei kuulu referenssiaineistoon
StreamProbRF_gini100ref_class	desimaali	Puroisuuden todennäköisyys, Random Forest mallinnus, päätöspuiden määrä/n_estimators=100, kriteeri='gini'. Todennäköisyys vaihtelee 0 (oja) ja 1 (puro) välillä. Välitulos, jonka ajossa käytetty aikaisempia parametrisarvoja. Arvoja voi käyttää validointiin. Käytä mieluummin StreamProbRF-saraketta mallin tuloksarvoina. Jos mallin kaikki ajot antavat samoja tuloksarvoja, tulos todennäköisesti pitää paikkansa, kuin jos tuloksarvojen välillä on paljon hajontaa.
StreamProbNN_BinCrossEntr_AdamStreamProbRF_gini100	desimaali	Puroisuuden todennäköisyys, Neural network mallinnus, model.compile(loss="binary_crossentropy", optimizer="adam", metrics="accuracy"). Todennäköisyys vaihtelee 0 (oja) ja 1 (puro) välillä. Välitulos, jonka ajossa käytetty aikaisempia parametrisarvoja. Arvoja voi käyttää validointiin. Käytä mieluummin StreamProbNN-saraketta mallin tuloksarvoina. Jos mallin kaikki ajot antavat samoja tuloksarvoja, tulos todennäköisesti pitää paikkansa, kuin jos tuloksarvojen välillä on paljon hajontaa.
predicted_class	desimaali	Mallinnusten keskiarvo (StreamProbRF_gini100, StreamProbNN_BinCrossEntr_Adam) Välitulos, jossa käytetty aikaisempia parametrisarvoja. Käytä mieluummin predicted_class_mean-saraketta mallin tuloksarvoina.

Muuttuja	Tyyppi	Kuvaus
Year	kokonaisluku, 32bit	Viivan lähtöaineiston vuosiluku. Mukana vain ne 2009 ja 2014 vuoden aineistot, jotka eivät leikkaa vuoden 2024 aineistoa.
min_distance_seq	desimaali	Kahden perättäisen solmupisteen minimietäisyys
max_distance_seq	desimaali	Kahden perättäisen solmupisteen maksimietäisyys
mean_distance_seq	desimaali	Kahden perättäisen solmupisteen keskietäisyys
stdev_distance_seq	desimaali	Kahden perättäisen solmupisteen etäisyyksien keskihajonta
StreamProbRF	desimaali	Puroisuuden todennäköisyys, Random Forest-mallinnus, päätöspuiden määrä/n_estimators=100, kriteeri='gini'. Todennäköisyys vaihtelee 0 (oja) ja 1 (puro) välillä.
StreamProbNN	desimaali	Puroisuuden todennäköisyys, Neural network-mallinnus, model.compile(loss="binary_crossentropy", optimizer="adam", metrics="accuracy"). Todennäköisyys vaihtelee 0 (oja) ja 1 (puro) välillä.
predicted_class_mean	desinmaali	Puromallinnusten keskiarvo (StreamProbRF, StreamProbNN). Todennäköisyys vaihtelee 0 (oja) ja 1 (puro) välillä.
unique_id	teksti	Yksilöivä id-tunnus
streamfilter	kokonaisluku, 32bit	Purosuodatus ojitusaineistoa varten (https://ckan.ymparisto.fi/dataset/soiden-ja-kivennaismaiden-ojitustilanne) Arvo 1= potentiaalinen puro, arvo 0 = potentiaalinen oja. Suodatus: "num_vertices" > 6 AND "sinuosity" > 1.02 AND "mean_distance_seq" < 10 AND ("StreamProbRF" > 0.6 OR "StreamProbNN" > 0.6)