



Prepared Petri Välisuo	Security	Date 16.4.2021	Revision A
---------------------------	----------	-------------------	---------------

Tekoäly ja luokittelu

1 Johdanto

Tekoälyn määritellään usein olevan koneen suorittamaa järkiperäistä päättelyä. Järkiperäisyydellä tarkoitetaan sitä, että päättely perustuu muodollisiin päättelysääntöihin.

Tämän osuuden tavoitteena on tutustua koneoppimiseen, joka on yksi tekoälyn osa-alueista. Koneoppimisessa algoritmi oppii datan perusteella tietyssä mielessä optimaaliset säännöt, joiden avulla data voidaan esimerkiksi luokitella halutulla tavalla.

Luokittelun tarkoituksena on määrittellä tai ennustaa mihin luokkaan tutkittava kohde ominaisuuksiensa puolesta kuuluu. Esimerkiksi kananmunapakkaamon lajittelurobotti erottelee pilaantuneet kananmunat tuoreista painon ja läpinäkyvyyden avulla. Se siis luokittelee kananmunat kahteen luokkaan: tuoreet ja pilaantuneet. Samalla tavoin jätteitä lajittelevan robotin pitää pystyä luokittelemaan jätteet esimerkiksi eri muovi- ja metallityyppeihin kierrätyksen tehostamiseksi. EU maksaa maataloustukia viljelyspinta-alojen mukaan, ja eri kasvien viljelyä valvotaan luokittelemalla maa-alueet kasvillisuuden perusteella satelliittikuvien perusteella. Luokittelualgoritmi pystyy siis määrittelemään kasvaako tietyllä peltolohkolla viljaa vai heinää tai onko se kesannolla.

Tässä työssä tutustutaan yksinkertaisen luokittimen käyttöön ja toteuttamiseen eri kurjenmiekkalajien erottelemiseksi toisistaan.

1. Ensimmäisessä vaiheessa luokitin toteutetaan Orange Data Mining ohjelmistolla. Orangen etuna on sen graafisuus, ja sen avulla voidaankin nopeasti tutkia dataa ja toteuttaa esimerkiksi luokitin ilman ohjelmointia.
2. Toisessa vaiheessa tutustutaan Pythonilla toteutettuun kurjenmieikka-aineiston luokittimeen, ja kokeillaan sen käyttöä. Lopussa on vielä muutama ohjelmointitehtävä, missä päästään kokeilemaan oman luokittelijan toteuttamista.



Prepared Petri Välisuo	Security	Date 16.4.2021	Revision A
---------------------------	----------	-------------------	---------------

2 Työn suorittaminen

2.1 Luokittaminen toteutus Orange Data Mining ohjelmistolla

2.1.1 Ohjelmiston asentaminen

Jos käytät lukion tietokonetta, kysy aluksi opettajaltasi, onko Orange data mining sovellus jo asennettuna lukion koneelle, jos ei ole, asenna se itse ohjeiden mukaan.

Orange Data Mining ohjelmisto voidaan asentaa yksinkertaisimmillaan lataamalla ZIP-tiedosto koneellesi ja purkamalla se haluamaasi hakemistoon. ZIP-tiedosto on melko iso (437 MT) ja se vie purettuna noin 1.2 GT levytilaa. Varmista että levytilaa on riittävästi ennen asennusta.

2.1.2 Orangen käynnistäminen

Tuplaklikkaa Orange–alihakemistosta Orange–nimistä linkkiä ja ohjelmisto käynnistyy. Kuva käyttöliittymästä on Kuvassa (1).

Orangeen käynnistyttyä näkyviin tulee tyhjä työtila ja vasemmassa laidassa on paletti erilaisia työkaluja joita voit raahata työtilaan. Käytä hetki aikaa ja tarkastele mitä työkaluja paletista löytyy ja miten ne on ryhmitelty. Työkalut voi tuoda työtilaan myös klikkaamalla työtilaa hiiren oikealla painikkeella ja etsimällä ne alasvetovalikosta. Valikossa on myös hakukenttä, josta työkalut löytyvät helposti nimen kirjoittamalla osa työkalun nimestä hakukenttään.

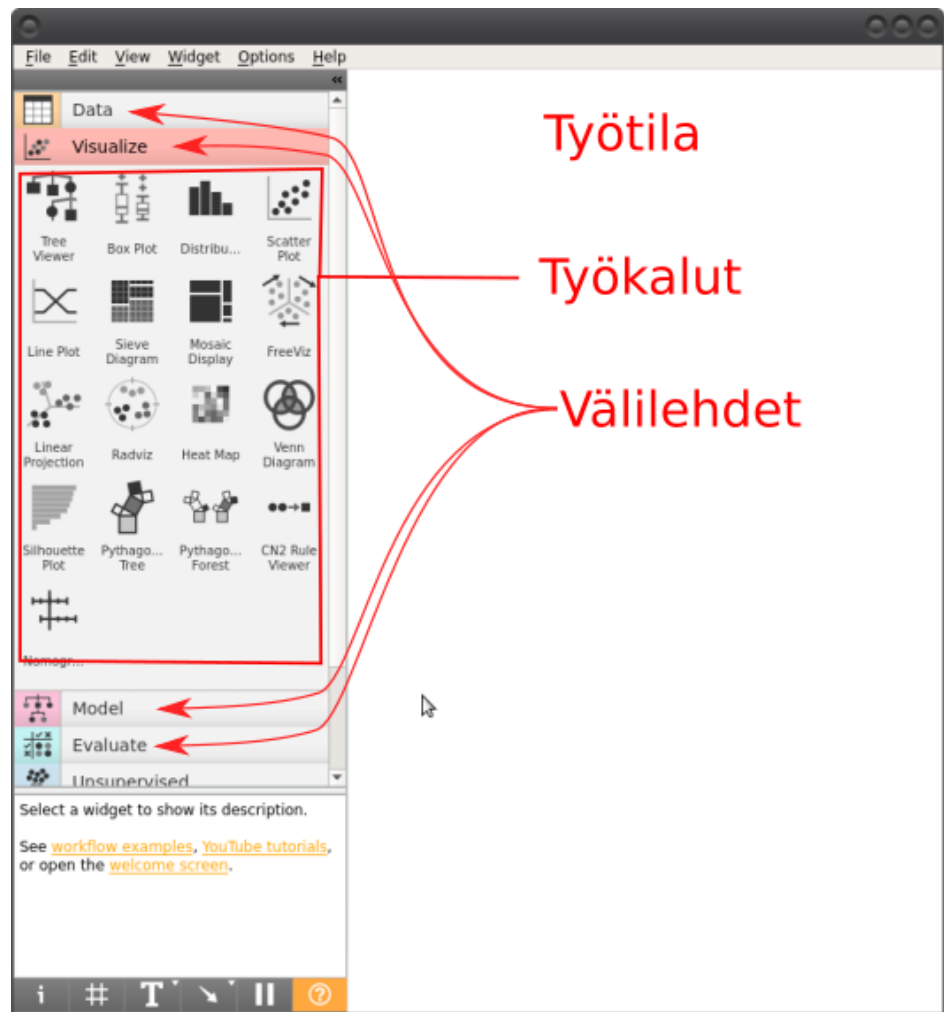
2.1.3 Aineiston lukeminen ja tarkastelu

Tuo aluksi työkalupaletin *Data*-välilehdeltä, *File* niminen työkalu työalueelle. Tuplaklikkaa *File*-työkalua ja valitse ponnahdusikkunan valikon yläreunasta tiedostoksi ”iris.tab”, jos sitä ei ole vielä valmiiksi valittu. Sen jälkeen voit sulkea ponnahdusikkunan.

Tuo tämän jälkeen työkalupaletin samalta *Data*-välilehdeltä ”*Data Table*” niminen työkalu työtilaan. Yhdistä nyt *File*, ja *Data Table* työkalut toisiinsa vetämällä hiirellä viiva *File*-työkalun oikeassa reunassa olevasta harmaasta kaariviivasta *Data-Table* työkalun vasemmassa reunassa olevaan kaariviivaan. Tämä yhdistää tiedostosta luettavan tiedon *Data Table*- työkalun sisääntuloon. Kun nyt tuplaklikkaat *Data Table* työkalun kuvaketta, aukeaa ponnahdusikkuna, josta voit tarkas-



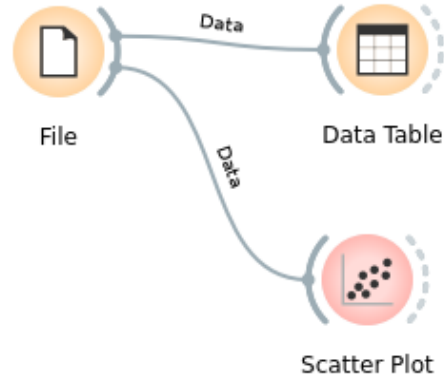
Prepared Petri Väლისuo	Security	Date 16.4.2021	Revision A
---------------------------	----------	-------------------	---------------



Kuva 1: Orangen käyttöliittymä



Prepared Petri Välisuo	Security	Date 16.4.2021	Revision A
---------------------------	----------	-------------------	---------------



Kuva 2: Aineiston luku ja tarkastelu

tella aineistoa taulukkomuodossa. Tutki dataa ja eri mahdollisuuksia järjestää data taulukossa ja katso miten voit pylväsdiagrammeilla visualisoida muuttujien suuruudet. Kun olet katsonut tarpeeksi, sulje poñnahdusikkuna.

Raahaa nyt työkalupaletin *Visualize*-välilehdeltä *Scatter-Plot* niminen työkalu työtilaan. Yhdistä tiedostosta saatava data *Scatter Plot*- työkaluun samoin kuin teit *Data Table*- työkalun kanssa ja tuplaklikkaa *Scatter Plot*- ikkuna auki. Nyt sinun pitäisi nähdä aineisto kahden muuttujan mukaisesti piirrettynä koordinaatistoon. Jokainen piste vastaa yhtä aineiston riviä, eli yhtä kukkaa. Valitse ikkunan vasemmasta reunassa olevasta *Color*-alasetoalistokista *C iris*, jolloin kukkia kuvapisteeet väritetään kukkatyyppin mukaisesti. X- ja Y-akselilla on valittuna yksi kukkaa kuvaavista muuttujista: Verholehden pituus ja leveys tai terälehten pituus ja leveys. Voit valita X-akselille valitun muuttujan vasemmalta ylhäältä alasetoalistokista. Samoin voit valita myös Y-akselilla olevan muuttujan. Kokeile eri vaihtoehtoja ja valitse se, missä eri kukkatyyppit erottuvat parhaiten toisistaan, eli missä eriväriset pisteet parhaiten erottuvat omiin ryhmiinsä.

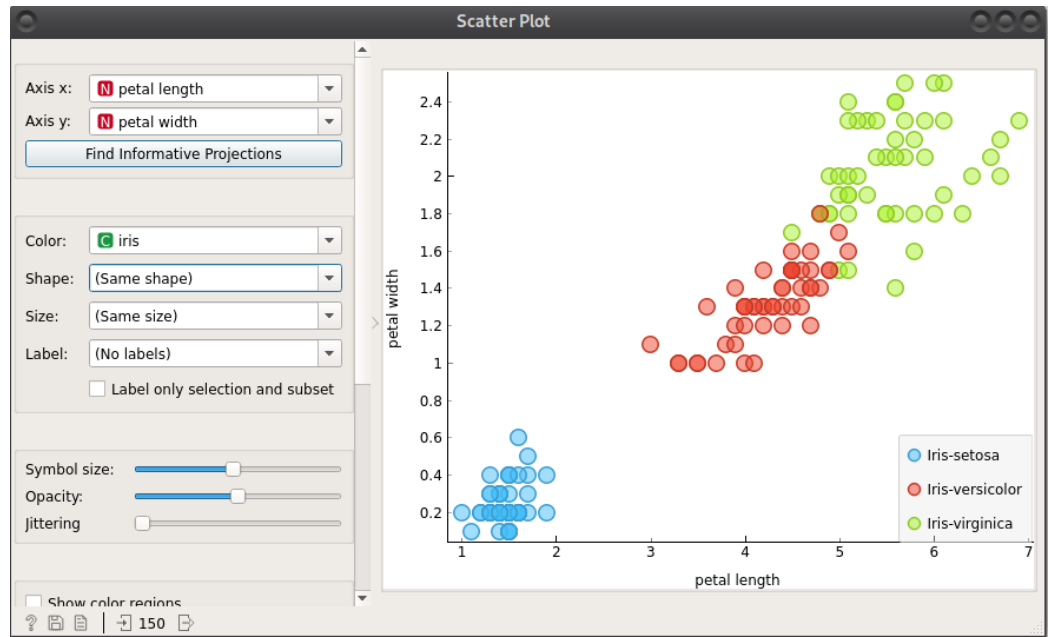
Työtilasi pitäisi nyt näyttää Kuvan (2) mukaiselta ja *Scatter plot* Kuvan (3) mukaiselta.

2.1.4 Luokittimen opettaminen

Lisää työtilaan päätöspuuokitin (*Tree*) raahaamalla se työkalupaletin *Model*-välilehdeltä työtilaan. Yhdistä sen sisääntuloon (harmaa kaari vasemmassa reunassa) tiedostosta tuleva aineisto samalla tavalla kuin teit aikaisemminkin.



Prepared Petri Välisuo	Security	Date 16.4.2021	Revision A
---------------------------	----------	-------------------	---------------



Kuva 3: Kurjenmiekkadatan kukkatyyppien mukaan väritettynä ja terälehtien pituuden ja leveyden mukaan kuvaajaan tulostettuna.

Lisää seuraavaksi työtilaan *Evaluate*-välilehdeltä ”*Test and Score*”- työkalu luokittimen testaamiseksi. Yhdistä sen sisääntuloon tiedostosta saatava aineisto kuten aikaisempiinkin työkaluihin. Yhdistä ”*Test and Score*”- työkalun vasempaan reunaan myös ”*Tree*”- työkalun oikeasta reunasta saatava päätöspuumalli. Sitten tuplaklikkaa ”*Test and Score*”- työkalua, ja tarkista luokittimen luokitustarkkuus ”*Precision*”. Avaa yhtäaikaan auki *Tree*- mallin ponnahtusikkuna ja voit kokeilla miten päätöspuun parametrin muuttaminen vaikuttaa luokittelutarkkuuteen. Palauta lopulta oletusarvot, jos vielä muistat mitä ne olivat. Jos et muista, poista puuluokitin (del-näppäimellä) ja raahaa uusi tilalle työkalupaletista.

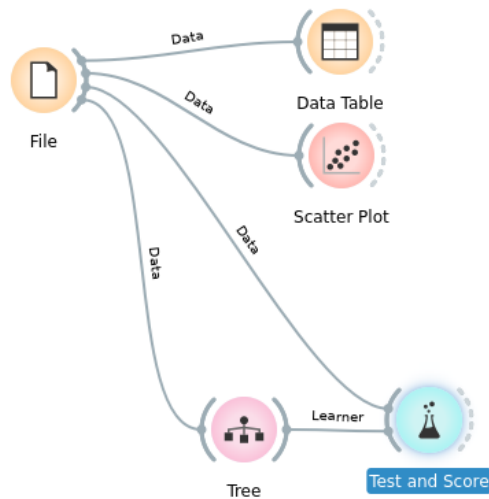
Työtilasi pitäisi nyt näyttää kutakuinkin Kuvan (4)kaltaiselta.

2.1.5 Luokittimen toiminnan tarkastelu

Tarkastellaan seuraavaksi, miten luokitin toimii. Raahaa *Visualize*-välilehdeltä ”*Tree Viewer*”- työkalu työtilaasi, ja kytke siihen vasemmalle puolelle *Tree*-työkalun oikea reuna ja. Kytke sitten ”*Tree Viewerin*” oikea puoli, eli ulostulo työtilalle aikaisemmin tuomasi ”*Scatter Plot*”-työkalun sisääntuloon (vasempaan reunaan). Työtilasi pitäisi nyt näyttää Kuvan (5) kaltaiselta.



Prepared Petri Välisuo	Security	Date 16.4.2021	Revision A
---------------------------	----------	-------------------	---------------



Kuva 4: Luokittimen opettaminen ja testaaminen

Tuplaklikkaa nyt auki sekä ”Tree Viewer”- että ”Scatter Plot”- työkalut ja tarkastele ”Tree Viewer”- työkalun avulla minkälaisilla päätössäännöillä eri kukkalajikkeet on luokiteltu. Kun klikkaat yhtä päätöspuun haaraa tai lehteä, siihen liittyvät pisteet ”Scatter Plot” työkalussa aktivoituvat.

1. Mitkä kukat tulivat väärin luokitelluiksi? Mistä se voi johtua?
2. Mikä laji on kyseessä jos sen terälehdien pituus (*petal length*) on 5 cm ja leveys (*petal width*) on 2 cm?

2.2 Luokitin Pythonilla

Nyt kun sekä luokitin, että ohjelmointi ovat tuttuja käsitteitä, niin tarkastellaan äsken Orangella toteutetun luokittimen toteuttamista Pythonilla. Avaa tämä työkirja, lue ohjeet, tarkastele luokittimen toteutusta ja suorita lopussa olevat tehtävät.

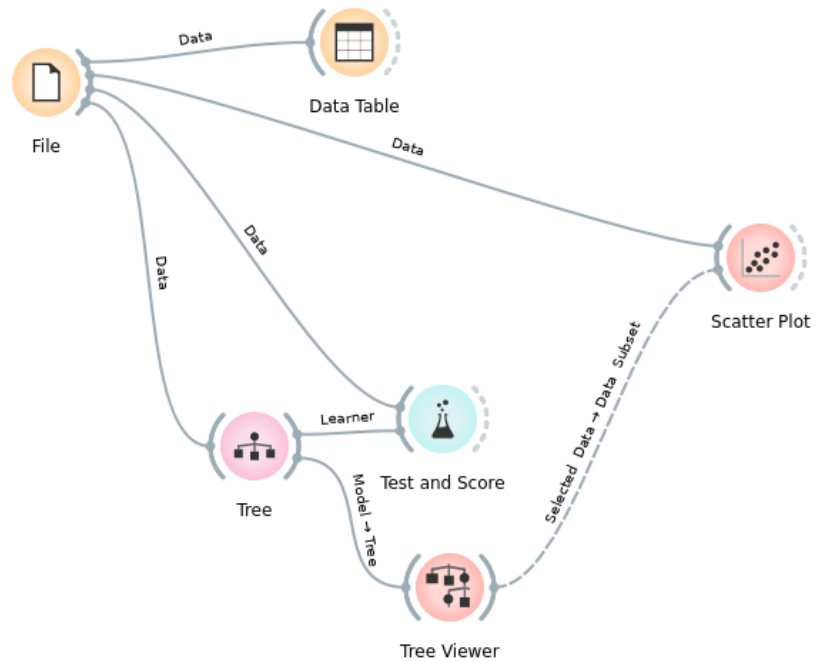
Huom! Tallenna kopio työkirjasta ”File/Save” valikon avulla itsellesi Google driveen, jotta voit myös tallentaa siihen tekemäsi muutokset.

2.3 Python perusteet

Ohjelmointi liittyvä vahvasti tekoälyyn. Jos aikaa jää, voidaan myös tutustua hiivenen Python ohjelmointiin



Prepared Petri Välisuo	Security	Date 16.4.2021	Revision A
---------------------------	----------	-------------------	---------------



Kuva 5: Päättöpuuluokittimen toiminnan tarkastelu

Avaa tämä työkirja, lue ohjeet ja suorita tehtävät.

Huom! Tallenna kopio työkirjasta ”File/Save” valikon avulla itsellesi Google driveen, jotta voit myös tallentaa siihen tekemäsi muutokset.