

POSTACADEMISCHE OPLEIDING

# MACHINE LEARNING

## VAN THEORIE TOT PRAKTIJK

19 maart 2024 – 18 juni 2024



UNIVERSITEIT  
GENT

Machine learning wordt gezien als de motor van de vierde industriële revolutie. De wendbaarheid van bedrijven moet maximaal zijn om te kunnen blijven overleven. De sleutel tot het overleven van deze nieuwe industriële revolutie is deze niet te ondergaan, maar in de bestuurderszetel te zitten door ervaring op te doen met machine learning.

Er wordt vaak gezegd dat data de nieuwe olie is, maar niet alle data is gelijk, noch staan grote hoeveelheden data garant voor grote waarde. Het is dankzij data analyse en machine learning dat waarde uit de verzamelde data kan gehaald worden. Machine learning wordt dan ook gezien als de motor van de vierde industriële revolutie. De wendbaarheid van bedrijven moet maximaal zijn om te kunnen blijven overleven. Het Internet of Things genereert continu data en er zijn momenteel al meer toestellen en machines die data genereren dan de gehele mensheid bij elkaar. Wie bovenop de data zit én hier machine learning succesvol op weet toe te passen, wordt de koploper en kan een hele sector op zijn kop zetten en domineren. Dit bewezen bedrijven zoals Amazon, Google, Netflix, Uber, en nog heel veel andere ondernemingen reeds met hun succesverhalen. De sleutel tot het overleven van deze nieuwe industriële revolutie is dan ook niet deze te ondergaan, maar in de bestuurderszetel te zitten door ervaring op te doen met machine learning.

Naast de theoretische achtergrond, wordt daarom in deze opleiding ook aandacht besteed aan het verwerven van praktisch inzicht via hands-on ervaring. Er wordt een overzicht gegeven van de verschillende machine learning principes en technieken, de valkuilen en de best-practices. Deze opleiding zal u naast een diepgaande theoretische basis ook de essentiële praktische know-how aanleveren die u zal kunnen gebruiken om zelf met machine learning aan de slag te gaan voor uw projecten.

In de cursus komen zowel clustering-, classificatie- als regressiemethoden aan bod, en worden deze toegepast op real-life data sets. We focussen in deze cursus vooral op de basisconcepten en -technieken, maar gaan ook dieper in op de toepassing ervan voor onder andere anomaliedetectie. De data extractietechnieken om deze data sets aan te leggen enerzijds en technieken om de data op te schonen en visualiseren anderzijds worden ook behandeld. In parallel met de verschillende lessen, kan je ook vrijblijvend deelnemen aan een diepgaander project waar je (alleen of in groep) alle tot dan toe geziene technieken kan combineren en gebruiken om een praktisch probleem uit de industrie op te lossen. Je werkt zelfstandig aan dit project, maar we voorzien ook 3 sessies om feedback te geven aan alle deelnemers van dit project.

## DOELPUBLIEK

De lessen zijn bedoeld voor iedereen die een goede professionele vertrouwdheid met informatica heeft en die graag praktisch, via hands-on sessies, aan de slag wil met machine learning. Deelnemers hebben een hogere opleiding in de informatica gevolgd of hebben een gelijkwaardige ervaring opgebouwd.

Deelnemers dienen programmeerervaring met Python of een aanverwante programmeertaal te hebben. Er wordt een inleidend labo "Introduction to Python" gegeven. Dit labo dient om de basisvaardigheden in programmeren in Python op te frissen (maar niet om Python vanaf nul aan te leren). Er wordt gewerkt met eigen laptop. Deze moet krachtig genoeg zijn (minimum 8GB RAM) en deelnemers moeten administratierechten hebben voor het installeren van de nodige programma's.

Het aantal deelnemers is beperkt tot 25.

## GETUIGSCHRIFT

U ontvangt een getuigschrift, indien u deelneemt aan de volledige opleiding en slaagt voor het bijhorende examen (schriftelijk, multiple choice). Het examen staat ingepland op 24 september 2024 om 16u30.

## WETENSCHAPPELIJKE COÖRDINATIE

**Prof. dr. ir. Sofie Van Hoecke**, Vakgroep Elektronica en Informatiesystemen, Faculteit Ingenieurswetenschappen en Architectuur, Universiteit Gent

## LESGEVERS

- **Prof. dr. ir. Sofie Van Hoecke**, IDLab, Vakgroep Elektronica en Informatiesystemen, Faculteit Ingenieurswetenschappen en Architectuur, Universiteit Gent
- **Prof. dr. ir. Jan Verwaeren**, Vakgroep Data-analyse en wiskundige modellering, Faculteit Bio-ingenieurswetenschappen, Universiteit Gent
- **Prof. dr. Willem Waegeman**, Vakgroep Data-analyse en wiskundige modellering, Faculteit Bio-ingenieurswetenschappen, Universiteit Gent

MEER INFO EN INSCHRIJVEN

[WWW.UGAIN.UGENT.BE/MACHINELEARNING](http://WWW.UGAIN.UGENT.BE/MACHINELEARNING)

# PROGRAMMA

## INTRODUCTIE MACHINE LEARNING

### Les 1

We starten met de verschillende categorieën van machine learning: unsupervised learning, supervised learning, reinforcement learning en eventuele hybride varianten. We leren het verschil tussen clustering, classificatie en regressie. We focussen ook op verschillende types data: structured data, image data, tijdreeksdata. In het tweede deel van deze les frissen we onze basisvaardigheden in programmeren in Python op.

### Les 2

We duiken dieper in op evaluatiemetrieken en data cleaning. We bespreken het verschil tussen test error rate en training error rate, en waarom het zo belangrijk is een train-, test- en validatieset te hebben. Cross-validatie zal ons helpen wanneer we geen grote beschikbare test set hebben. Tenslotte bespreken we ook enkele praktische tips om een goede testset op te bouwen. We gaan ook direct aan de slag met Nearest Neighbor methodes, een zeer eenvoudige classificatiemethode die verrassend goed werkt bij veel problemen. Deze les sluiten we af met een labo omtrent ETL en kNN.

Data: 19 en 26 maart 2024  
Lesgever: Sofie Van Hoecke

## DATAVISUALISATIE

Datavisualisatie is een belangrijk onderdeel van elk machine learningproject om meer te weten te komen over de beschikbare data en om eerste patronen te identificeren. We behandelen basis plotting bibliotheken zoals plotly, seaborn en matplotlib, en leren hoe we hoog-dimensionale data kunnen visualiseren. Ook clustering komt in deze sessie aan bod, zowel als finale taak, of gebruikt tijdens de data exploratie en preprocessing. Feature aggregatie en principale componenten analyse, t-SNE en UMAP zullen gebruikt worden wanneer we een grote hoeveelheid gegevens willen beschrijven met een kleiner aantal features.

Datum: 16 april 2024  
Lesgever: Jan Verwaeren

## REGRESSIE

Lineaire regressie is een zeer eenvoudige, maar zeer nuttige methode om een kwantitatieve uitkomst te voorspellen. We bespreken lineaire regressie als basistechniek en gaan ook in op de tekortkomingen en nieuwere, complexere benaderingen.

Datum: 30 april 2024  
Lesgever: Willem Waegeman

## CLASSIFICATIE

Er bestaan heel wat classificatietechnieken, en daarom bespreken we hier enkele van de meest gebruikte en/of eenvoudige classificatiemethodes naast k-Nearest Neighbours die reeds in de eerste les gezien werd, zijnde Naive Bayes en logistische regressie.

Datum: 7 mei 2024  
Lesgever: Sofie Van Hoecke

## MODEL SELECTIE EN REGULARISATIE

In deze les gaan we dieper in op het finetunen van lineaire modellen via regularisatie. We bespreken eerst feature selection als een algemene methode om de performantie te verhogen en inzicht te verwerven in een machine learning model. Vervolgens bespreken we twee lineaire methoden die de basis vormen voor complexe methoden die later in de cursus aan bod komen: ridge regressie en lasso. Beiden zijn nauw met elkaar verwant, maar ze hebben enkele belangrijke verschillen. In het laatste deel van de les bespreken we hoe de besproken regularisatiemethoden ook toegepast worden op niet-lineaire modellen en in een classificatiecontext.

Datum: 14 mei 2024  
Lesgever: Willem Waegeman

## ENSEMBLE

Ensemble-methoden profiteren van de voordelen van beslissingsbomen, terwijl ze de neiging tot overfitting verminderen. We leren hoe bagging (bootstrap-aggregatie) of boosting kan helpen om de variantie en bias te verminderen.

Datum: 28 mei 2024  
Lesgever: Jan Verwaeren

## NEURALE NETWERKEN

Neurale netwerken kunnen niet ontbreken vanwege de vooruitgang die hier de afgelopen jaren geboekt is. Neurale netwerken modelleren de data met behulp van kunstmatige neuronen. Ook deze techniek kan gebruikt worden voor zowel classificatie als regressie. Ook transfer learning en incremental learning komen kort aan bod.

Datum: 4 juni 2024  
Lesgever: Sofie Van Hoecke

## ANOMALIEDTECTIE

Anomaliedetectie houdt zich bezig met het vinden van data die afwijken van normaal gedrag en speelt dan ook een grote rol in de industrie, met toepassingen zoals kwaliteitscontrole, machinemonitoring voor voorspellend onderhoud, tot spamdetectie en bewaking. We bespreken en doen hands-on ervaring op met verschillende anomaliedetectietechnieken, zowel unsupervised, semi-supervised, als supervised methodes, en methodes specifiek voor tijdsreeksdata.

Datum: 11 juni 2024  
Lesgever: Sofie Van Hoecke

## KAGGLE PROJECT

### Sessie 1

We starten met de data-analyse (statistieken en visualisatie) om inzicht in de data van het project te krijgen. Daarna gaan we over tot data cleaning en preparatie, alsook wat eerste (lineaire) baseline modellen voor het project.

### Sessie 2

Daarna gaan we verder met het finetunen van onze lineaire modellen, het opstellen van tree-based modellen en/of neurale netwerken om te zien of deze onze baseline modellen overtreffen. We kijken ook om een goede en betrouwbare evaluatie op te zetten.

### Sessie 3

Nadien kan er worden overgestapt naar het opstellen van een ensemble van alle voorgaande gecreëerde modellen om de laatste prestatiewinsten te halen en hopelijk hoog op de Kaggle leaderboard te eindigen.

Data: 23 april, 21 mei en 18 juni 2024  
Lesgever: Sofie Van Hoecke

## PRAKTISCH

### Prijs

De deelnameprijs bedraagt **2.400 euro**.

Deelnameprijs omvat lesgeld, hand-outs, frisdranken, koffie en broodjes.

Betaling geschiedt na ontvangst van de factuur. Alle facturen zijn betaalbaar dertig dagen na dagtekening. Alle vermelde bedragen zijn vrij van BTW.

### Korting

- Indien minstens één deelnemer van een bedrijf inschrijft voor de volledige opleiding wordt voor alle bijkomende gelijktijdige inschrijvingen van hetzelfde bedrijf een korting van 20% verleend. Facturatie geschiedt dan d.m.v. een gezamenlijke factuur.
- Aangepaste prijzen voor personeel van UGent
- Kortingen zijn niet cumuleerbaar.

### Annulering

Raadpleeg onze annulatievoorwaarden op [www.ugain.ugent.be/annulatievoorwaarden](http://www.ugain.ugent.be/annulatievoorwaarden)

### KMO-portefeuille

Universiteit Gent aanvaardt betalingen via de KMO-portefeuille: [www.kmo-portefeuille.be](http://www.kmo-portefeuille.be); gebruik autorisatiecode DV.0103194.

### Opleidingsverlof

Erkenningsaanvraag voor opleidingsverlof (VOV) werd aangevraagd.

### Tijdstip en locatie

De lessen (behalve de sessies van het kaggle project) worden **on campus** gegeven van 16u30 tot 20u, in 2 delen, gescheiden door een broodjesmaaltijd. Deze vinden plaats aan de Universiteit Gent, UGent Academie voor Ingenieurs, Technologiepark 60, 9052 Zwijnaarde.

Tijdens iedere lesavond wordt er 1,5 uur theorie gegeven en wordt er na de pauze 1,5 uur oefeningen voorzien.

De Kaggle-sessies gaan **online** door, van 16u30 tot 20u, waarbij elke deelnemer (of team van deelnemers) een slot toegekend zal krijgen om het project te bespreken, feedback te krijgen en next steps te definiëren.

Het examen vindt plaats op 24 september 2024 om 16u30.

Data onder voorbehoud van wijzigingen om onvoorziene omstandigheden.

### Laptop

Er wordt gewerkt met eigen laptop. Deze moet krachtig genoeg zijn (minimum 8GB RAM) en deelnemers moeten administratierechten hebben voor het installeren van de nodige programma's.

### Organisatie

**Universiteit Gent**  
UGain (UGent Academie voor Ingenieurs)  
Technologiepark 60  
9052 Zwijnaarde  
09 264 55 82  
[ugain@ugent.be](mailto:ugain@ugent.be) - [www.ugain.ugent.be](http://www.ugain.ugent.be)

### Met de steun van VAIA



MEER INFO EN INSCHRIJVEN

[WWW.UGAIN.UGENT.BE/MACHINELEARNING](http://WWW.UGAIN.UGENT.BE/MACHINELEARNING)