

POSTACADEMISCHE OPLEIDING

MACHINE LEARNING

VAN THEORIE TOT PRAKTIJK

6 september 2022 – 20 december 2022



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 niet deze 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, aan de hand van 3 modules, 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. Naast de basisconcepten, gaan we ook dieper in op meer geavanceerde machine learning technieken zoals anomaliedetectie, time series analyse en aanbevelingssystemen. 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 modules, 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 per module telkens ook 1 sessie 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 hebben programmeerervaring met Python of een aanverwante programmeertaal. Er kan een vrijblijvend labo "Introduction to Python" gevolgd worden. 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 is gepland op 24 januari 2023 om 17u30.

WETENSCHAPPELIJKE COÖRDINATIE

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

Prof. dr. Willem Waegeman, Vakgroep Data-analyse en wiskundige modellering, Faculteit Bio-ingenieurswetenschappen, Universiteit Gent

LESGEVERS

- **Dr. ir. Jim Clauwaert**, Vakgroep Data-analyse en wiskundige modellering, Faculteit Bio-ingenieurswetenschappen, Universiteit Gent
- **Prof. dr. ir. Toon De Pessemier**, Waves, Vakgroep Informatietechnologie, Faculteit Ingenieurswetenschappen en Architectuur, Universiteit Gent
- **Prof. dr. Stijn Luca**, Vakgroep Data-analyse en wiskundige modellering, Faculteit Bio-ingenieurswetenschappen, Universiteit Gent
- **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

PROGRAMMA

0. ZELFSTUDIE PYTHON

Vrijblijvende online module om basisvaardigheden in programmeren in Python op te frissen

1. INTRODUCTIE MACHINE LEARNING

In deze eerste module introduceren we de basisterminologie en concepten achter machine learning.

Sessie 1: Intro

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, en gaan dieper in op de basis evaluatiemetrieken en data cleaning. We gaan ook direct aan de slag met Nearest Neighbor methodes, een zeer eenvoudige classificatiemethode die verrassend goed werkt bij veel problemen.

Sessie 2: Data visualisatie

Datavisualisatie is een belangrijk onderdeel van elk machine learning-project om meer te weten te komen over de beschikbare data en eerste patronen te identificeren. We behandelen basis plotting bibliotheken zoals plotly, plotnine 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 zal gebruikt worden wanneer we een grote hoeveelheid gegevens willen beschrijven met een kleiner aantal features.

Sessie 3: Regressie

Lineaire regressie is een zeer eenvoudige maar zeer nuttige methode om een kwantitatieve uitkomst te voorspellen. Naast het bespreken van lineaire regressie als basistechniek, gaan we ook in op de tekortkomingen en nieuwere, complexere benaderingen waarbij lineaire regressie aan de basis ligt. Ook gradient descent en feature engineering voor niet-vectoriële data komen in deze sessie aan bod.

Sessie 4: Lineaire classificatie

Wanneer de output kwalitatief of categorisch is in plaats van kwantitatief, dienen we classificatietechnieken te gebruiken in plaats van regressie. Er bestaan heel wat technieken, en daarom bespreken we hier de twee meest gebruikte classificatiemethodes naast k-Nearest Neighbours die reeds in de eerste les gezien werd, zijnde logistische regressie en lineaire discriminantanalyse.

Sessie 5: Kaggle project

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.

2. BASISCONCEPTEN

We zoomen verder in op de basisconcepten van machine learning, van beslissingsbomen en ensembles tot neurale netwerken.

Sessie 1: Evaluatie van machine learning methoden

In deze les gaan we dieper in op het evalueren van machine learning algoritmen, en de mogelijke valkuilen die hierbij kunnen optreden. 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, en we bespreken ook kort bootstrapping als een alternatief voor cross-validatie. Tenslotte bespreken we ook enkele praktische tips om een goede testset op te bouwen.

Sessie 2: Model selectie & regularisatie

In deze les gaan we dieper in op het fine-tunen 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.

Sessie 3: Tree-based classificatie

Beslissingsbomen zijn zeer populaire technieken die zowel voor classificatie als regressie gebruikt kunnen worden. Ze zijn interpreteerbaar en kunnen ook gebruikt worden voor feature selection. 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

Sessie 4: Neurale netwerken

(Diepe) neurale netwerken kunnen hier 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 aan bod.

Sessie 5: Kaggle project

In deze les gaan we verder met het fine-tunen 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..

3. GEAVANCEERDE METHODEN

In deze laatste module gaan we dieper in op meer geavanceerde machine learning technieken.

Sessie 1: Convolutionele Neurale Netwerken

In deze les gaan we dieper in op gespecialiseerde neurale netwerken, meer specifiek convolutionele neurale netwerken. De introductie van convolutionele neurale netwerken zorgden voor een enorme verbetering in taken zoals beeldherkenning, maar hebben de laatste jaren ook veel success geboekt in andere velden. Deze vooruitgang ging verder gepaard met allerhande architecturale veranderingen aan het netwerk die het optimaliseren van zeer diepe modellen mogelijk maakten.

Sessie 2: Anomaliedetectie

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.

Sessie 3: Aanbevelingssystemen en collaboratieve filtering

Er is een stijgend belang voor aanbevelingssystemen in industriële toepassingen, gedreven door de beschikbaarheid van verschillende soorten gebruikersgegevens. We geven een actueel overzicht van technieken voor aanbevelingssystemen die in een industriële omgeving worden gebruikt. Ook collaboratieve filtering als specifieke case komt aan bod om gebruikersvoorkeuren te voorspellen als lineaire combinatie van andere gebruikers hun voorkeuren.

Sessie 4: Recurrente Neurale Netwerken

Er zijn ook veel bedrijfstoeepassingen van tijdsreeksvoorspelling, zoals het voorspellen van aandelenkoersen, verkoop, of het weer, tot het voorspellen van de nog resterende tijd alvorens een component zal falen. We zullen verschillende modellen voor tijdsreeksvoorspelling zien, zoals Recurrente Neurale Netwerken (RNN), en leren wat de voor- en nadelen van zijn.

Sessie 5: Kaggle project

In deze laatste sessie halen we alles uit de kast en stappen we over naar geavanceerde machine learning modellen tot misschien zelfs 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.

PRAKTISCH

Prijs

De deelnameprijs bedraagt **2.500 euro**.

Deelnameprijs omvat lesgeld, hand-outs, het handboek 'An introduction to statistical learning', 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.

Lesmateriaal

Naast de hand-outs ontvangen de deelnemers ook het handboek 'An introduction to statistical learning'. Dit handboek is inbegrepen in de deelnameprijs.

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.

Annulering

Raadpleeg onze annulatievoorwaarden op www.ugain.ugent.be/annulatievoorwaarden

Tijdstip en locatie

- De lessen (sessie 1 tot 4 per module) worden on campus gegeven **van 17u30 tot 21u**, 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 (sessie 5 per module) kunnen vrijblijvend gevolgd worden en vinden online plaats van 17u30 tot 21u, 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 januari 2023 om 17u30.
- Data onder voorbehoud van wijzigingen om onvoorziene omstandigheden.

KMO-portefeuille

Universiteit Gent aanvaardt betalingen via de KMO-portefeuille: www.kmo-portefeuille.be; gebruik autorisatiecode DV.0103194.

Opleidingsverlof

Deze opleiding is erkend in het kader van VOV.

0. ZELFSTUDIE PYTHON

1. INTRODUCTIE MACHINE LEARNING

6 september 2022	Sessie 1: Intro Sofie Van Hoecke
13 september 2022	Sessie 2: Data visualisatie Jan Verwaeren
20 september 2022	Sessie 3: Regressie Stijn Luca
27 september 2022	Sessie 4: Lineaire classificatie Stijn Luca
4 oktober 2022	Sessie 5: Kaggle project (online) Sofie Van Hoecke & Willem Waegeman

2. BASISCONCEPTEN

11 oktober 2022	Sessie 1: Evaluatie van machine learning methoden Willem Waegeman
18 oktober 2022	Sessie 2: Model selectie & regularisatie Willem Waegeman
25 oktober 2022	Sessie 3: Tree-based classificatie Jan Verwaeren
8 november 2022	Sessie 4: Neurale netwerken Sofie Van Hoecke
15 november 2022	Sessie 5: Kaggle project (online) Sofie Van Hoecke & Willem Waegeman

3. GEAVANCEERDE METHODEN

22 november 2022	Sessie 1: Convolutionele Neurale Netwerken Jim Clauwaert
29 november 2022	Sessie 2: Anomaliedetectie Sofie Van Hoecke
6 december 2022	Sessie 3: Aanbevelingssystemen en collaboratieve filtering Toon De Pessemier
13 december 2022	Sessie 4: Recurrente Neurale Netwerken Jim Clauwaert
20 december 2022	Sessie 5: Kaggle project (online) Sofie Van Hoecke & Willem Waegeman

Organisatie

Universiteit Gent
UGain (UGent Academie voor Ingenieurs)
Technologiepark 60
9052 Zwijnaarde
09 264 55 82
ugain@ugent.be - www.ugain.ugent.be

Met de steun van: