

De toekomst van machine learning

vraaggesprek met Daniel Situnayake

C. J. Abate (Content Director, Elektor)

Daniel Situnayake is een creatieve ingenieur die van mening is dat embedded machine learning (ML) een technologie is die je maar eens in je leven tegenkomt. Hier spreekt hij over het potentieel van ML en introduceert hij TinyML en enkele ideale toepassingen. We spreken ook kort over zijn ervaring als ontwikkelaar bij Google (figuur 1).

Figuur 1. Daniel Situnayake
(Founding TinyML Engineer, Edge Impulse).



Van de universiteit naar het bedrijfsleven

Abate: Laten we beginnen met uw achtergrond. Wanneer bent u voor het eerst geïnteresseerd geraakt in embedded ML? Was het op de Birmingham City University (BCU), waar u een BSc in computernetwerken en beveiliging behaalde? Of was het nadat u de universiteit verliet?

Situnayake: Ik was niet van plan me met embedded machine learning te gaan bezighouden. In feite was het een lange weg naar waar ik nu ben, met veel ervaringen die me in de richting van dit terrein duwden! Na mijn afstuderen aan de BCU werd me gevraagd om bij de universiteit te blijven werken, les te geven en namens de universiteit als consultant op te treden. Ik richtte me op het onderwerp van automatische identificatie- en data capture-technieken, dat zijn manieren waarop een computer fysieke objecten kan herkennen en verwerken – alles van biometrie, zoals gezichtsherkenning, tot RFID, barcodes en smartcards. Dit was mijn eerste kennismaking met twee dingen: computer vision-algoritmen, die destijds niet gebaseerd waren op machinaal leren, en embedded systemen. Een paar banen later in mijn carrière, na een acquisitie, werkte ik in het ontluikende data science-team bij een consumentenbank. Hier kwam ik voor het eerst in aanraking met de tools en workflows van *big data*. Het was een heel nieuw onderwerp voor mij en ik vond het fascinerend.

Toen we met Tiny Farms begonnen, zag ik voor het eerst een kans om deze terreinen te combineren, door gebruik te maken van data science-tools om embedded sensordata te interpreteren en

te gebruiken in de agrarische technologie. Maar het was bij Google, waar ik Pete Warden (**figuur 2**) en Andy Selle van het TensorFlow Lite-team ontmoette, dat ik echt het potentieel van *deep learning* op kleine apparaten begreep, en daar ben ik me op gaan concentreren.

Abate: Wanneer bent u naar Californië verhuisd? Was het voor een specifieke carrièrestap?

Situnayake: Ik verhuisde in 2009 naar Californië, in het jaar nadat de wereldeconomie instortte. Ik had de jaren daarvoor een groot aantal mentale en fysieke tests ondergaan met als doel aangenomen te worden bij de Britse luchtmacht. Ik wilde met Apache-helikopters vliegen! Ik zou in mei 2009 beginnen met de officiersopleiding, dus ik ging op een afscheidsreis door de VS met mijn beste vriend. Toen we in Californië waren, ontmoette ik een meisje. Een paar maanden later, kort na het eerste deel van mijn opleiding in Sandhurst, zag ik dat een leven als gentleman-piloot niets voor mij was en dat een ontspannen leven in Californië beter bij me paste. Ik boekte een vlucht, trok in bij het meisje dat ik net had ontmoet, en zes maanden later zijn we getrouwd!

Dat huwelijk heeft geen stand gehouden, maar sindsdien woon ik hier. Toen ik de overstap maakte van de academische wereld naar het bedrijfsleven, sloot ik me aan bij een AI-startup in Los Angeles en verhuisde ik naar de Bay Area, net op tijd voor de technologie-boom van 2010.

Insectenboerderij

Abate: U bent mede-oprichter van Tiny Farms, dat u op LinkedIn omschrijft als “America’s first insect farming technology company”. Hoe is dat bedrijf tot stand gekomen? En hoe staat het ervoor?

Situnayake: Ook dit verhaal begint met een reis door de VS. Twee goede vrienden dachten na over een project om het probleem van de voedselvoorziening aan te pakken. Tijdens een trip in de vrije natuur overnachtten ze in een hut bij een meer dat omgeven was door grote, luidruchtige sprinkhanen. In een opwelling vroegen ze zich af of die beesten eetbaar waren. Na wat onderzoek leerden ze dat eetbare insecten eigenlijk heel belangrijk zijn voor de wereldwijde voedselvoorziening, maar dat de technieken om ze in gevangenschap te kweken nauwelijks waren onderzocht.

Ze begonnen te experimenteren en vroegen mij of ik met hen een bedrijf wilde beginnen. Ik zag veel mogelijkheden om de technieken die ik had geleerd rond automatische identificatie en data capture toe te passen op het automatiseren van het kweken van insecten, en in het gebruik van datawetenschappen om de iteratieve ontwikkeling van nieuwe systemen te begeleiden.

Ik ben er vol overgave in gedoken! In het begin was het een parttime-project, maar uiteindelijk hebben we geld ingezameld en een insectenkwekerij op industriële schaal gebouwd in San Leandro in Californië. Het was intens, spannend en behoorlijk stressig om miljoenen krekels in leven te houden en intussen bezig te zijn met klanten, investeerders en onderzoek. Zo’n vijf jaar later



Figuur 2. Daniel Situnayake en Pete Warden signeren hun TinyML-boek.

helpen. Dat leidt tot allerlei spannende projecten, maar het kan ook een beetje overweldigend zijn: hoe bepaal je waar je je voor wilt inzetten, als er zoveel mogelijkheden zijn?

Dat maakt het ook wel lastig. Omdat er altijd zoveel aan de hand is, is er een kans dat je eideloo vergadert en afgeleid wordt. Het eerste

“

Google heeft een verbazingwekkend vrije bedrijfscultuur: je kunt nagenoeg aan alles werken wat je maar wilt, zolang het je product vooruit helpt.

”

besloot ik dat het tijd was om de handdoek in de ring te gooien. Mijn medeoprichters gingen nog een paar jaar door, maar uiteindelijk bleek het te moeilijk om een product te vinden dat goed aansloot op de markt en hielden zij er ook mee op. Tiny Farms leeft voort als een open source-repository met ontwerpen en gegevens die in die zeven jaar zijn verzameld [1].

Werken bij Google

Abate: U hebt bijna drie jaar bij Google gewerkt. Voordat we ingaan op de inhoud van uw werk, ben ik ook wel nieuwsgierig naar hoe het was om bij zo’n indrukwekkend bedrijf te komen.

Situnayake: Na vijf jaar opbouwen van een agrarisch bedrijf, waar ik meer tijd besteedde aan bouwwerkzaamheden dan aan software, wilde ik weer in de softwareontwikkeling duiken, en er was geen bedrijf dat ik technisch gezien meer bewonderde dan Google.

Al gauw wist ik niet wat me overkwam. Ik had nog nooit bij een groot technisch bedrijf gewerkt. Met je Google-badge heb je toegang tot alle faciliteiten in elk Google-kantoor, waar ook ter wereld. Dus het voelde alsof ik een burger was geworden van een vreemd nieuw soort land, met eigen openbare diensten, infrastructuur en overheid.

Google heeft een verbazingwekkend vrije bedrijfscultuur: je kunt nagenoeg aan alles werken wat je maar wilt, zolang het je product vooruit helpt. Als je een cool idee hebt, word je aangemoedigd om dat uit te voeren en andere mensen aan boord te halen om te

jaar dat je bij Google werkt, ontwikkel je een soort intern filter om te voorkomen dat je overweldigd wordt door alles wat er gebeurt. Het is een sprookje dat iedereen die bij Google werkt een soort genie is. Bij Google vind je hetzelfde scala aan vaardigheden als in de rest van de technische industrie. Ik ben nooit iemand geweest die hielden op een voetstuk plaatst, maar het is een fijn gevoel als je verschillende topmensen in je vakgebied ontmoet en je ontdekt dat het gewone mensen als jij en ik zijn.

Abate: U hebt bij Google gewerkt als Developer Advocate voor TensorFlow Lite. Wat is TensorFlow Lite? En hoe hebt u de ontwikkelaars opgeleid? Door middel van meetups? Online-cursussen?

Situnayake: TensorFlow is de naam voor Google’s ecosysteem van open-source tools voor training, evaluatie en het inzetten van deep learning-modellen. Het omvat alles van de high-level Python-code die wordt gebruikt om modelarchitecturen te definiëren tot en met de low-level code die wordt gebruikt om die modellen uit te voeren op verschillende soorten processoren.

TensorFlow Lite is de subset van de TensorFlow-tools om modellen op zogenaamde ‘edge devices’ in te zetten; dat betreft eigenlijk alles dat kleiner is dan een personal computer: telefoons, microcontrollers en meer. De TensorFlow Lite-tools kunnen met een deep learning-model werken en dat optimaliseren, zodat het beter geschikt wordt om op dit soort apparaten te draaien. Het omvat ook de sterk geoptimaliseerde code die op de apparaten zelf draait.

Mijn favoriete onderdeel van TensorFlow Lite is de gestripte, supersnelle variant die is ontworpen om deep learning-modellen te draaien op kleine, goedkope, low-power microcontrollers. Voorafgaand aan de lancering van *TensorFlow Lite for Microcontrollers* moesten ontwikkelaars hun eigen low-level code schrijven voor het uitvoeren van ML-modellen op embedded apparaten. Dat maakte de drempel om toegang te krijgen tot on-device machineren ongelooflijk hoog. De introductie van TensorFlow haalde veel obstakels uit de weg, zodat ineens iedereen met enige ervaring in embedded systemen modellen kon draaien. Het was heel interessant om te zien hoe het veld van TinyML tot leven kwam.

Het was mijn rol om het TensorFlow Lite-team te helpen het ecosysteem van de ontwikkelaar te begrijpen en er verbinding mee te maken. Alle ontwikkelaars zijn verschillend, dus het is belangrijk om verschillende manieren te creëren om ze erbij te betrekken. We hebben alles gedaan wat je kunt voorstellen, van persoonlijke ontmoetingen en door Google gefinancierde conferenties tot het bouwen van gemakkelijk te begrijpen voorbeeldcode en het schrijven van het TinyML-boek!

Abate: *Waarom bent u uiteindelijk bij Google weggegaan? Was dat een moeilijke beslissing?*

Situnayake: Bij het werken met de ontwikkelaars van TensorFlow heb ik gezien dat er zelfs bij de beste ML-tools en -bibliotheken enorm veel kennis nodig is om ze te kunnen gebruiken. Er zijn uitstekende cursussen en materialen, maar de tijd die je moet investeren maakt dat machine learning buiten het bereik van veel drukbezette technici blijft. Opleidingsmodellen zijn net zo goed een kunst als een wetenschap, en het kost jaren om het gevoel ervoor te ontwikkelen en de beste methoden onder de knie te krijgen. Bij TinyML is het nog moeilijker, omdat het een compleet nieuw vakgebied is en de beperkingen niet goed bekend zijn.

Toen ik hoorde over Edge Impulse, het bedrijf waar ik heen ging toen ik Google verliet, was ik erg onder de indruk. Nog voor de officiële lancering was het de oprichters gelukt om een product te bouwen dat elke ontwikkelaar in staat stelt om deep learning-modellen te trainen die op embedded hardware kunnen draaien. Dankzij mijn ervaring met de *developer community* zag ik in, dat zij een aantal van de lastigste problemen in het veld hadden opgelost, waardoor veel pijnpunten bij het werken met ML-technieken verdwenen. Embedded ML is volgens mij een baanbrekende technologie. Het gaat de wereld om ons heen veranderen op manieren die we nog niet kunnen overzien. Edge Impulse is opgericht om deze technologie beschikbaar te maken voor alle engineers, zodat die hun uiteenlopende perspectieven en grote kennis van hun vakgebied kunnen gebruiken om complexe problemen op te lossen.

Zodra ik dat in de gaten had, besloot ik om me erbij aan te sluiten. Het was best jammer om mijn vrienden bij Google te moeten verlaten, maar TinyML is een klein wereldje, dus ik werk met veel van hen nog steeds nauw samen!

Tiny Machine Learning (TinyML)

Abate: *Wat moeten professionals en serieuze makers over TinyML weten? En kunt u wat voorbeelden van ultra-low-power ML-applicaties noemen?*

Situnayake: De wereld telt inmiddels waar je maar kijkt miljarden sensoren, van huishoudelijke apparaten tot slimme fabrieken en de voertuigen die we gebruiken. Al die sensordata vormt een

gouden kans voor ontwikkelaars om apparaten te bouwen die de wereld om zich heen begrijpen en die intelligente beslissingen nemen om mensen te helpen.

In het verleden was de enige manier om gebruik te maken van deze gegevens ze via het internet naar een krachtige computer te sturen, die de getallen kan kraken, de gegevens interpreteren en actie ondernemen op basis van de resultaten. Voor sommige toepassingen werkt dat prima. Het is bijvoorbeeld OK als de gegevens van een weerstation een paar seconden nodig hebben om bij een server te komen. Maar er zijn veel toepassingen waar dit niet werkt. Een van mijn favoriete voorbeelden hiervan is de camera-val. Wetenschappers gebruiken die om dieren in het wild in de gaten te houden, vaak op afgelegen plaatsen waar geen internetverbinding beschikbaar is. Als een dier voorbij komt, wordt de bewegingssensor van de camera geactiveerd en wordt er een foto gemaakt. De onderzoekers komen regelmatig langs om een geheugenkaartje met dierenfoto's op te halen.

Het probleem is dat de bewegingssensor door allerlei dingen kan worden geactiveerd, zoals bladeren die in de wind bewegen, of door diersoorten waarin de onderzoekers niet geïnteresseerd zijn. Het duurt lang om alle foto's te doorzoeken, en de geheugenkaart raakt snel vol, wat betekent dat mensen vaak naar buiten moeten om die om te wisselen. Zelfs als er een internetverbinding beschikbaar zou zijn, zou het extra stroomverbruik betekenen dat de batterijen van de camera snel leeg raken en vervangen moeten worden.

Door het toevoegen van een TinyML-model dat getraind is om de juiste diersoort te herkennen, kan worden voorkomen dat foto's van dwarsrelende bladeren worden gemaakt. Dat betekent dat zo'n val langer autonoom kan werken, wat de onderzoekers tijd en geld bespaart.

Er zijn vier belangrijke randvoorwaarden die een toepassing ideaal maken voor TinyML. Als een applicatie een beperkte connectiviteit heeft, een lage tolerantie voor latentie, een beperkte energiebron of strenge privacybeperkingen, kan TinyML mogelijk helpen omdat het niet nodig is om gegevens vanaf het apparaat te versturen.

Het is belangrijk om te weten dat alleen bepaalde problemen kunnen worden opgelost door middel van machine learning. Het moeten problemen zijn waar een beetje vaagheid acceptabel is, omdat ML zelden absolute antwoorden kan geven. Het is heel geschikt voor het maken van dierenfoto's, waarbij foutjes acceptabel zijn, maar het is misschien niet betrouwbaar genoeg in sommige veiligheidskritische toepassingen.

Abate: *U bent co-auteur van het boek "TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers". Omdat u Arduino behandelt, denk ik dat de TinyML-technologie toegankelijk is voor zowel amateurs als professionals. Wat kunnen engineers en innovatoren nog meer doen om met TinyML aan de slag te gaan?*

Situnayake: Zoals altijd is zelf enkele projecten bouwen de beste manier om te leren! Een goede plek om te beginnen vormen enkele van de tutorials die we aanbieden in de Edge Impulse-documentatie: bewegingsherkenning met een versnellingsopnemer [2]; het herkennen van geluiden [3]; en het toevoegen van zicht aan uw sensoren [4].

We hebben Edge Impulse gebouwd zodat u uw mobiele telefoon kunt gebruiken om gegevens te verzamelen en modellen te testen, zodat u aan de slag kunt zonder dat u speciale hardware nodig hebt. Als u eenmaal vertrouwd bent met ML-methodiek, kun u uw

favoriete ontwikkelkaart pakken en beginnen te hacken! Ik zou de Arduino Nano 33 BLE Sense aanbevelen als een goed systeem om te beginnen met TinyML. Deze heeft een snelle maar zuinige Cortex M-4F processor, en genoeg RAM voor interessante modellen. Hij wordt volledig ondersteund in Edge Impulse.

Edge Impulse

Abate: U bent in januari 2020 bij Edge Impulse in San Jose gekomen. Hoe is dat gegaan?

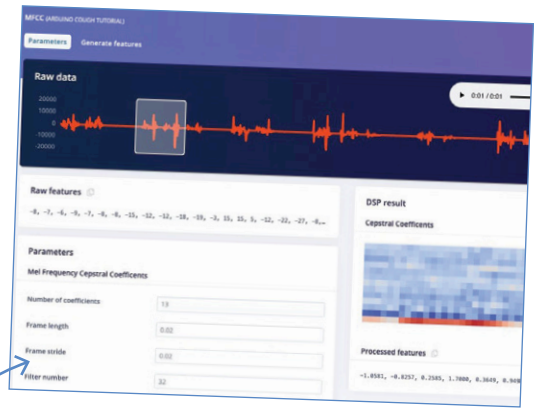
Situnayake: Ik had de CEO, Zach Shelby, leren kennen toen ik nog bij Google werkte. Voorheen was hij VP bij ARM, dat zeer nauw samenwerkt met het TensorFlow Lite-team. Toen Zach aankondigde dat hij ARM ging verlaten om een TinyML-bedrijf op te richten, trok dat mijn aandacht en toen ze hun eerste vacature plaatsten, moest ik wel solliciteren!

Abate: Hoe ziet uw werkdag eruit? Schrijft u de hele dag code? Geeft u les? Doet u business development?

Situnayake: Een van de fijnste dingen bij Edge Impulse is dat ik weer in de eerste plaats een engineer ben. Ik breng het grootste deel van mijn tijd door met hacken. Ik heb nog wel contact met klanten en dat is erg nuttig, omdat ik daardoor weet waar ik mijn technische inspanningen op moet richten. Ik ben ook erg actief in de TinyML-community, dus ik geef regelmatig lezingen en help met het organiseren van TinyML-meetups. Meestal heb ik 's morgens wel een paar vergaderingen, maar daarna kan ik de rest van de dag lekker programmeren!

Abate: Ik zie dat het Edge Impulse team in mei is gestart met het COVID-19-gerelateerde project "Hoestdetectie met TinyML op Arduino" (figuur 3). Het project is meer dan 5.400 keer bekeken. Hoe is de respons vanuit de engineering community geweest? Wat is de status van het project?

Situnayake: Hoewel Edge Impulse nog maar een paar maanden bestaat, hebben we al een fantastische community met verschillende achtergronden. Toen COVID-19 toesloeg, was een van de leden van onze community, Kartik Thakore, datawetenschapper en biomedisch engineer, geïnteresseerd in het bouwen van een aantal projecten om te helpen. We dachten dat dit een mooie kans zou zijn om te laten zien wat er mogelijk is met TinyML, dus we stuurden hem wat hardware en hij stelde een project samen om hoest te detecteren met behulp van een TinyML-model. Het resultaat was dit project [5], waarvan we hopen dat het de makers inspireert om soortgelijke uitdagingen aan te gaan!



Figuur 3. Hoestdetectie met TinyML op Arduino (Bron: Edge Impulse).

TinyML-projecten

Abate: Zijn er nog meer TinyML-projecten die u onder de aandacht van onze lezers wilt brengen?

Situnayake: Bij Edge Impulse geloven we sterk in de positieve invloed van technologie, dat is duidelijk te zien bij enkele van mijn favoriete projecten. Zo traint Smart Parks bijvoorbeeld TinyML-modellen om wilde olifanten op te sporen, zodat de dorpelingen gewaarschuwd kunnen worden als ze op weg zijn naar de stad. Ik ben ook heel blij met mensen die met computers onverwachte dingen weten te doen waar je nooit aan zou hebben gedacht. Benjamin Cabé creëerde bijvoorbeeld een prachtig project dat gebruik maakt van een gassensor om verschillende merken alcoholische dranken te herkennen! Als u het wilt reproduceren, is er een geweldige tutorial van Seeed Studio. 

200423-04

Vragen of opmerkingen?

Heeft u vragen of opmerkingen naar aanleiding van dit artikel? Stuur een e-mail naar de auteur of naar de redactie van Elektor, via redactie@elektor.com.

Een bijdrage van

De vragen werden gesteld door: **C. J. Abate**
Vertaling: **Evelien Snel**
Layout: **Harmen Heida**

WEBLINKS

- [1] Open Tiny Farms: <https://www.opentinyfarms.com/>
- [2] Edge Impulse, "Continuous Motion Recognition": <https://docs.edgeimpulse.com/docs/continuous-motion-recognition>
- [3] Edge Impulse, "Recognize Sounds from Audio": <https://docs.edgeimpulse.com/docs/audio-classification>
- [4] Edge Impulse, "Adding Sight to Your Sensors": <https://docs.edgeimpulse.com/docs/image-classification>
- [5] Edge Impulse, "Cough Detection with TinyML on Arduino": <http://bit.ly/edge-cough-detection-project>