

# Machine Learning Pipeline Automation

Amar Kisoensingh

16 februari 2021



---

# Voorwoord

---

## **Samenvatting**

---

## Summary

---

---

---

# Woordenlijst

**cloud computing-platform** 12

**machine learning** is het ontwikkelen van algoritmes en technieken waarmee computers kunnen leren. 12

**machine learning pipeline** is een pijplijn waarin gedefiniëerd staat hoe data stroomt. 12, 13

---

## Lijst van figuren



---

# Inhoudsopgave

<b>Voorwoord</b>	<b>3</b>
Samenvatting . . . . .	4
Summary . . . . .	5
<b>Afkortingen</b>	<b>6</b>
<b>Woordenlijst</b>	<b>7</b>
<b>Lijst van figuren</b>	<b>8</b>
<b>1 Inleiding</b>	<b>11</b>
1.1 Het bedrijf: NGTI . . . . .	12
1.1.1 Klanten van NGTI . . . . .	12
1.1.2 Tools die worden gebruikt . . . . .	12
1.2 Opdracht . . . . .	12
1.3 Leeswijzer . . . . .	13
<b>2 Probleemanalyse</b>	<b>14</b>
2.1 Probleemdefinitie . . . . .	15
2.2 Bestaand oplossingen . . . . .	15
2.3 Doelstelling . . . . .	15
2.4 Hoofd- en deelvragen . . . . .	15
<b>3 Onderzoeksmethodes</b>	<b>16</b>
3.1 Kwalitatieve onderzoeksmethode . . . . .	17
3.2 Kwantitatieve onderzoeksmethode . . . . .	17
3.3 Onderzoeksmethodes bij de hoofd- en deelvragen . . . . .	17
<b>4 Requirementsanalyse</b>	<b>18</b>
4.1 Stakeholders . . . . .	19
4.2 Requirements . . . . .	19
<b>5 Hoe werkt Machine Learning?</b>	<b>20</b>
5.1 Soorten Machine Learning . . . . .	21
5.2 Wachtijden voorspellen . . . . .	21
5.3 Conclusie . . . . .	21
<b>6 Hoe werkt een Machine Learning pipeline</b>	<b>22</b>
6.1 Literatuur . . . . .	23
6.2 POC . . . . .	23
6.3 Conclusie . . . . .	23
<b>7 Cloud computing platformen en lokale frameworks</b>	<b>24</b>
7.1 Literatuur . . . . .	25
7.2 Conclusie . . . . .	25

---

<b>8</b>	<b>Architecturale ontwerp van de oplossing</b>	<b>26</b>
8.1	Literatuur . . . . .	27
8.2	Design . . . . .	27
8.3	Conclusie . . . . .	27
<b>9</b>	<b>Oplossing</b>	<b>28</b>
9.1	Scope definiëren . . . . .	29
9.2	Research techstack . . . . .	29
9.3	Wireframe . . . . .	29
9.4	Mockup . . . . .	29
9.5	POC . . . . .	29
9.6	Conclusie . . . . .	29
<b>10</b>	<b>Conclusie</b>	<b>30</b>
<b>11</b>	<b>Aanbeveling</b>	<b>31</b>
<b>12</b>	<b>Discussie</b>	<b>32</b>
<b>13</b>	<b>Reflectie</b>	<b>33</b>
	<b>Bijlagen</b>	<b>34</b>

# 1 Inleiding

---

## 1.1 Het bedrijf: NGTI

NGTI is mobile app development bureau gevestigd in Rotterdam, Nederland. Zij maakt hoogwaardige mobiele applicaties voor native, hybrid en webgebruik.

### 1.1.1 Klanten van NGTI

### 1.1.2 Tools die worden gebruikt

Om productief te zijn gebruikt NGTI een aantal tools en programma's om producten te maken en te communiceren met zowel collega's als klanten.

#### Slack

Interne communicatie gaat via Slack. Het programma faciliteert collega's om elkaar met een lage instap te benaderen en berichten die voor het hele bedrijf gelden te versturen. Ook zijn er 'channels' beschikbaar over specifieke onderwerpen, zoals: *#dev*, *#ios* en *#test-automation*.

#### Google Workspace

Met Google Workspace kunnen bestanden en documenten gemaakt, opgeslagen en gedeeld worden. Omdat dit via een browser kan, hoeven werknemers geen software te installeren. NGTI gebruikt het ook om collaboratief en parallel te werken aan hetzelfde document.

#### Zoom

Voorheen wordt Zoom alleen gebruikt om te videobellen met collega's en interviewees. In de tijd van het pandemie is Zoom echter een belangrijke speler geworden om effectief samen te werken. Meetings zoals introducties van nieuwe collega's of demo's van producten worden online gehouden.

## 1.2 Opdracht

NGTI heeft voorzien dat ze haar applicaties 'slimmer' moet maken door machine learning in te zetten. Niet alleen zorgt dit voor een verhoogde *user experience*, maar geeft NGTI ook een voorsprong op haar concurrenten.

Om machine learning toe te passen is het raadzaam om een machine learning pipeline op te zetten. Een pipeline is een gestructureerde werkwijze om een model te trainen. Het opzetten van de pipeline en een competente model trainen is tijdrovend en vereist kennis in het domein. Vaak worden modellen getraind op een cloud computing-platform zoals Azure, AWS of Google Cloud. Het probleem met platformen zoals deze is dat het ontzettend lastig is om te wisselen van platform.

---

De opdracht bestaat uit twee onderdelen:

1. onderzoek naar het automatiseren van het opzetten van een machine learning pipeline om het laagdrempelig en minder tijdrovend te maken.
2. onderzoek naar het maken van een platform agnostische oplossing

Hierbij zal een PoC gemaakt worden om aan te tonen of het haalbaar is. Een diepere duik in het probleem is te vinden in hoofdstuk 2.

## **1.3 Leeswijzer**

## 2 Probleemanalyse

- 
- 2.1 Probleemdefinitie**
  - 2.2 Bestaand oplossingen**
  - 2.3 Doelstelling**
  - 2.4 Hoofd- en deelvragen**

## 3 Onderzoeksmethodes



- 
- 3.1 Kwalitatieve onderzoeksmethode**
  - 3.2 Kwantitatieve onderzoeksmethode**
  - 3.3 Onderzoeksmethodes bij de hoofd- en deelvragen**

## 4 Requirementsanalyse

---

## **4.1 Stakeholders**

## **4.2 Requirements**

## 5 Hoe werkt Machine Learning?

---

**5.1 Soorten Machine Learning**

**5.2 Wachttijden voorspellen**

**5.3 Conclusie**

## 6 Hoe werkt een Machine Learning pipeline

---

**6.1 Literatuur**

**6.2 POC**

**6.3 Conclusie**

## 7 Cloud computing platformen en lokale frameworks



---

**7.1 Literatuur**

**7.2 Conclusie**

## 8 Architecturale ontwerp van de oplossing

---

**8.1 Literatuur**

**8.2 Design**

**8.3 Conclusie**

## 9 Oplossing

- 
- 9.1 Scope definiëren**
  - 9.2 Research techstack**
  - 9.3 Wireframe**
  - 9.4 Mockup**
  - 9.5 POC**
  - 9.6 Conclusie**

## 10 Conclusie

## 11 Aanbeveling

## 12 Discussie



## 13 Reflectie

## Bijlagen