

Sophie Bakker | Data Scientist met passie voor voorspellende modellen

Herengracht 123, Amsterdam, Noord-Holland, Nederland, 1011 AB

☎ +31 6 12345678 • ✉ sophie.bakker@datavak.nl • 🌐 <https://www.sophiebakker.nl>
🌐 [@sophiebakker](https://www.sophiebakker.nl) • 📺 [@sophiebakker](https://www.sophiebakker.nl)

Basis Informatie

Data Scientist met ruim 6 jaar ervaring in het ontwikkelen en implementeren van datagedreven oplossingen voor complexe vraagstukken.

- Ontwerpt en implementeert voorspellende modellen voor klantverloop, vraagvoorspelling en fraudedetectie.
- Combineert sterke analytische vaardigheden met een pragmatische, bedrijfsgerichte aanpak.
- Ervaren met Python, R, SQL en cloudplatformen zoals AWS en Google Cloud.
- Begeleidt junior datascientists en draagt bij aan de datastrategie van de organisatie.

Opleidingen

Universiteit van Amsterdam

Master, Data Science, Score: 8.2

<https://www.uva.nl>

sep 2017–jul 2019

- Afstudeeronderzoek naar causale inferentie in observationele studies, beoordeeld met een 9.
- Specialisatie in voorspellende modellen en datagedreven besluitvorming.

Cursussen: Machine Learning, Statistical Learning, Big Data Technologies, Deep Learning, Data Visualisatie, Ethiek en Data-beleid

Universiteit Utrecht

Bachelor, Kunstmatige Intelligentie, Score: 7.6

<https://www.uu.nl>

sep 2014–jul 2017

- Brede basis in AI-technieken, statistiek en softwareontwikkeling.
- Minor Ondernemerschap gevolgd aan de Universiteit Utrecht.

Cursussen: Logica en Redeneren, Algoritmiek, Datastructuren, Probabilistische Modellen, Mens-Machine Interactie, Cognitie-wetenschap

Werkervaring

Nova Analytics B.V.

Senior Data Scientist

<https://www.nova-analytics.nl>

jan 2022–Heden

- Geeft leiding aan de ontwikkeling van schaalbare machine-learning-pipelines voor klanten in de financiële sector.
- Verbeterde het churnmodel met 15% nauwkeurigheid door feature-engineering en modeloptimalisatie.
- Implementeerde MLOps-praktijken zoals CI/CD en modelmonitoring, wat de time-to-market van modellen met 40% verkortte.
- Adviseert het management over datastrategie en bouwt aan een sterk data science-team.

Trefwoorden: Python, Machine Learning, MLOps, Strategie, Teamlead

Zicht Data BV

Data Scientist

<https://www.zichtdata.nl>

sep 2019–dec 2021

- Analyseerde grote datasets om logistieke processen efficiënter te maken.
- Ontwikkelde realtime dashboards met Python, Streamlit en Tableau voor het monitoren van KPI's.
- Werkte nauw samen met productowners en domeinexperts om databehoeften te vertalen naar analyses.
- Voerde A/B-testen en statistische analyses uit ter ondersteuning van productbeslissingen.

Trefwoorden: Python, SQL, Data-analyse, Tableau, A/B-testen

Datazicht Labs

Data Analyst (stage)

<https://www.datazichtlabs.nl>

feb 2019–jul 2019

- Ondersteunde het team bij het opschonen, samenvoegen en analyseren van data voor beleidsonderzoek.
- Realiseerde interactieve visualisaties met R en R Shiny.
- Droeg bij aan datakwaliteitsrapportages en verbeterde de metadata van datasets.

Trefwoorden: R, Data Cleaning, Visualisatie, Rapportage

Talen

Nederlands: Moedertaal of tweetalige vaardigheid

Engels: Volledige professionele vaardigheid

Duits: Beperkte werkvaardigheid

Trefwoorden: Moedertaal

Trefwoorden: Cambridge C2, Zakelijk Engels

Trefwoorden: Basisvaardigheden

Vaardigheden

Machine Learning: Expert

Trefwoorden: Python, scikit-learn, TensorFlow, PyTorch, XGBoost, LightGBM

Statistiek: Geavanceerd

Trefwoorden: R, Hypothesetoetsing, Tiddreksen, A/B-testen, Causale Inferentie

Data Engineering: Intermediair

Trefwoorden: SQL, Apache Spark, Airflow, Docker, dbt, AWS

Data Visualisatie: Geavanceerd

Trefwoorden: Tableau, Power BI, matplotlib, seaborn, ggplot2, Plotly, D3.js

Datamanagement: Intermediair

Trefwoorden: Data governance, Datakwaliteit, Metadata, AVG/GDPR

Onderscheidingen

Master Scriptieprijs Data Science

Nederlandse Vereniging voor Data Science

nov 2019

Bekroond voor de beste afstudeerscriptie op het gebied van causale inferentie met observationele data.

KLM Datathon - Best Model

KLM

jun 2021

Team behaalde de eerste plaats met een voorspellingsmodel voor bagageafhandelingsvertragingen.

Certificaten

AWS Certified Machine Learning - Specialty

Amazon Web Services

<https://aws.amazon.com/certification/>

mei 2022

Google Professional Data Engineer

Google

<https://cloud.google.com/certification/data-engineer>

nov 2020

Publicaties

Tijdschrift voor Toegepaste Data Science

<https://www.tijdschriftdata.nl/artikel>

Een vergelijking van gradient boosting technieken voor vraagvoorspelling in de retailsector

Onderzoekt de prestaties van XGBoost, LightGBM en CatBoost op retaildata en geeft praktische aanbevelingen voor het toepassen van deze technieken.

Projecten

Energieprognose voor slimme wijken

Het project combineert tijdreeksanalyse met deep learning-modellen (LSTM) en biedt een API voor het ontsluiten van de voorspellingen.

Open-source project dat energieverbruik op buurtniveau voorspelt op basis van historische data en weersvoorspellingen. **Tref-**

woorden: Python, Tiddreksen, LSTM, Energietransitie

Fraudedetectiesysteem voor betaalverkeer

Implementeerde een realtime fraudedetectiepijplijn met Apache Kafka voor streaming en scikit-learn voor de detectiemodellen.

Eindproject van de masteropleiding: realtime fraudedetectiesysteem voor transacties. **Trefwoorden:** Fraudedetectie, Kafka, scikit-learn, Realtime

Interesses

Wielrennen: Toertochten, Amstel Gold Race

Schaken: Clubcompetitie, Strategie

Lezen: Sciencefiction, Vakliteratuur

Koken: Vegetarisch, Wereldkeuken

Vrijwilligerswerk

Data for Good Nederland

<https://www.dataforgoodnl.nl>

Data Scientist (vrijwilliger)

jan 2020-Heden

- Verzorgt data-analyses voor non-profitorganisaties, waaronder de optimalisatie van voedselbankdistributie.
- Ondersteunt vrijwilligers met het uitvoeren van dataprojecten.