



AUTOPRIJS VOORSPELLEN MET AZURE ML

Monday, October 31, 2022





AGENDA

- Introductie
- Data science overzicht
- Azure ML
- Autoprijs voorspellen
- Next steps

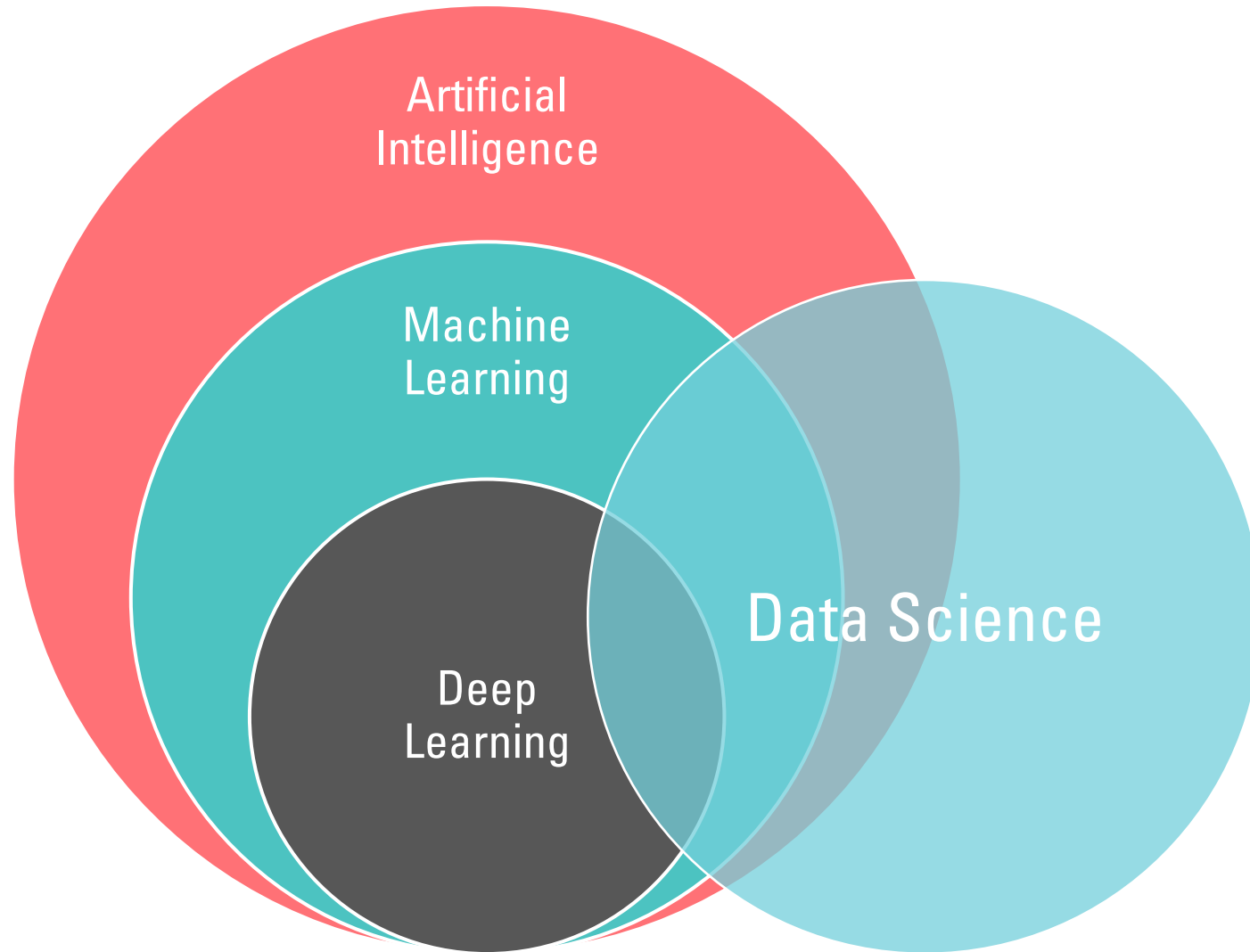
INTRODUCTIE



NIELS JONKER



DATA SCIENCE INTRODUCTIE



DOPPELGANGERS



OPSPOREN VAN STROPERS DOOR ANALYSE VAN DRONE BEELDEN



COMPLETED

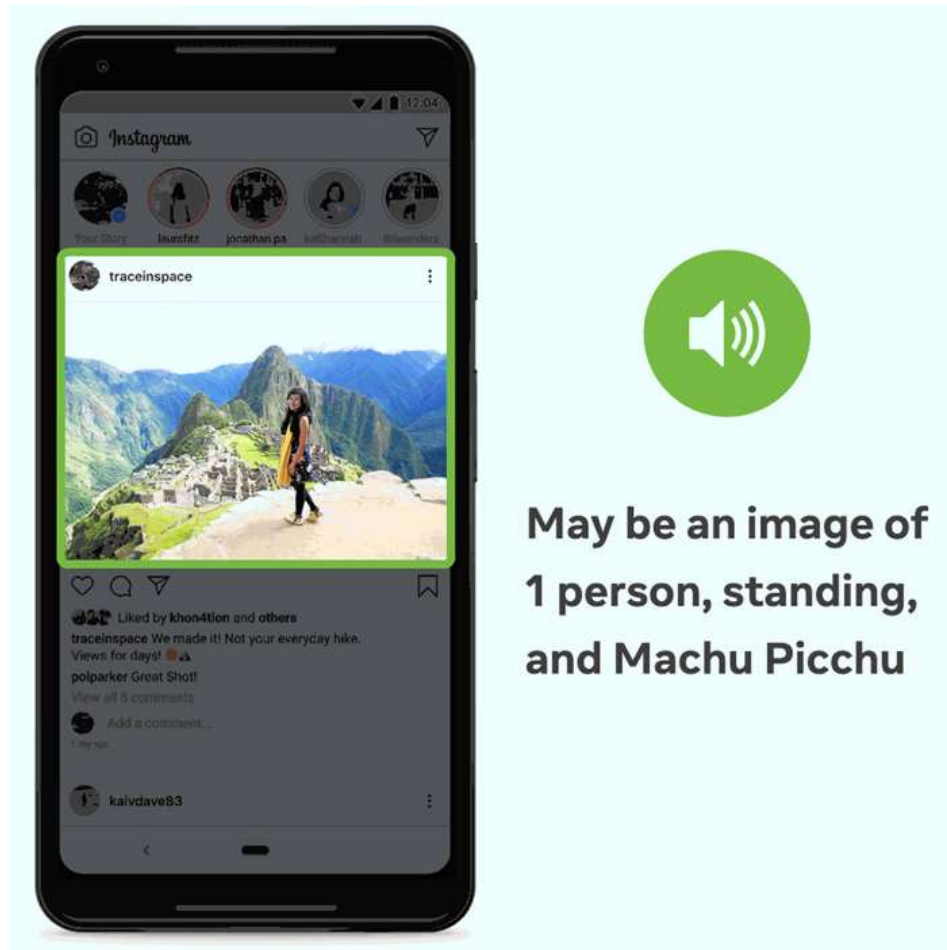
March 18, 2022

AI for Wildlife Challenge 3

AI for Wildlife 3 is bringing poacher-detecting ML models into production on a wildlife protection drone. In the third challenge of the AI for Wildlife series we'll be focusing on the elusive concept of MLOps.

[View challenge](#) [View more](#)

BEELD HERKENNING VOOR MENSEN MET SLECHT ZICHT



TEXT HERKENNING OM FORMULIEREN TE SCANNEN

Hack Day user feedback form

Your feedback helps us improve our events!

Name: [REDACTED] Position: [REDACTED]

Company: [REDACTED]

Phone: [REDACTED]

Twitter: [REDACTED]

Company: [REDACTED]

Position: [REDACTED]

Tags: [REDACTED]

1. Did you attend previous for this? ☐ yes ☐ no

2. How do you feel about this event? ☐ great ☐ good ☐ average ☐ bad ☐ terrible

3. Was the length of the event? ☐ too short ☐ just right ☐ too long

4. Overall value of the event? ☐ excellent ☐ good ☐ average ☐ bad ☐ terrible

5. How relevant was this to your work? ☐ extremely ☐ very ☐ somewhat ☐ not at all

6. How effective was the speaker? ☐ excellent ☐ good ☐ average ☐ bad ☐ terrible

7. Would you come again? ☐ yes ☐ maybe ☐ no

8. What was your main takeaway from this session?

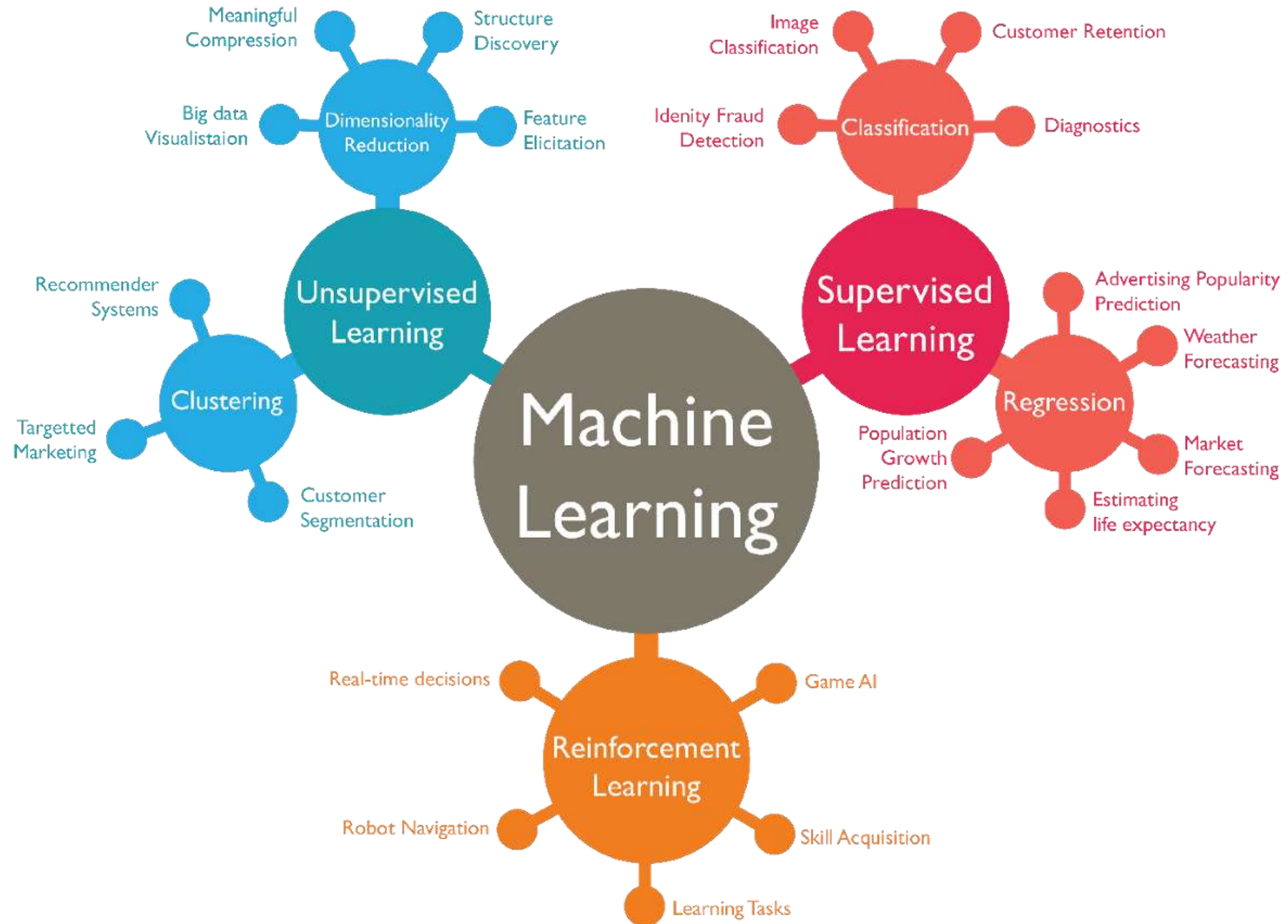
9. What do you think about the event?

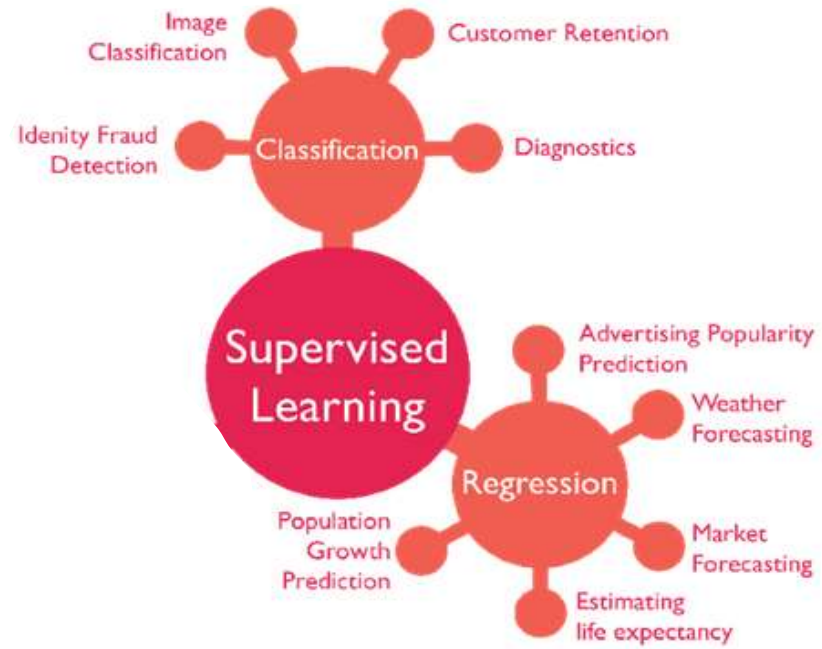
10. Do you have any tips for the speakers or organizers?

11. Are there any other topics you're interested in?

12. Would you like us to contact you further about our training or consulting?

MACHINE LEARNING BASICS





TRAIN TEST SPLIT

VOLLEDIGE DATASET

symboling	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	length	width	height	curb-weight	engine-type	num-of-cylinders	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price
2	toyota	gas	std	two	hardtop	rwd	front	98.4	176.2	65.6	52	2536	ohc	four	146	mpfi	3.62	3.5	9.3	116	4800	24	30	9639
1	dodge	gas	std	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6377
-1	volvo	gas	turbo	four	wagon	rwd	front	104.3	188.8	67.2	57.5	3157	ohc	four	130	mpfi	3.62	3.15	7.5	162	5100	17	22	18950
0	honda	gas	std	four	sedan	fwd	front	96.5	175.4	65.2	54.1	2465	ohc	four	110	mpfi	3.15	3.58	9	101	5800	24	28	12945
-1	toyota	gas	std	four	hatchback	fwd	front	102.4	175.6	66.5	53.9	2458	ohc	four	122	mpfi	3.31	3.54	8.7	92	4200	27	32	11248
-1	toyota	gas	std	four	hatchback	fwd	front	102.4	175.6	66.5	53.9	2414	ohc	four	122	mpfi	3.31	3.54	8.7	92	4200	27	32	9988
0	honda	gas	std	two	hatchback	fwd	front	96.5	167.5	65.2	53.3	2236	ohc	four	110	1bbl	3.15	3.58	9	86	5800	27	33	7895
0	toyota	gas	std	four	sedan	fwd	front	95.7	166.3	64.4	53	2094	ohc	four	98	2bbl	3.19	3.03	9	70	4800	38	47	7738
0	honda	gas	std	four	sedan	fwd	front	96.5	175.4	62.5	54.1	2372	ohc	four	110	1bbl	3.15	3.58	9	86	5800	27	33	10295
2	volkswagen	diesel	std	two	sedan	fwd	front	97.3	171.7	65.5	55.7	2261	ohc	four	97	idi	3.01	3.4	23	52	4800	37	46	7775

TRAIN DATA



symboling	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	length	width	height	curb-weight	engine-type	num-of-cylinders	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price
2	toyota	gas	std	two	hardtop	rwd	front	98.4	176.2	65.6	52	2536	ohc	four	146	mpfi	3.62	3.5	9.3	116	4800	24	30	9639
1	dodge	gas	std	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.4	68	5500	31	38	6377
-1	volvo	gas	turbo	four	wagon	rwd	front	104.3	188.8	67.2	57.5	3157	ohc	four	130	mpfi	3.62	3.15	7.5	162	5100	17	22	18950
0	honda	gas	std	four	sedan	fwd	front	96.5	175.4	65.2	54.1	2465	ohc	four	110	mpfi	3.15	3.58	9	101	5800	24	28	12945
-1	toyota	gas	std	four	hatchback	fwd	front	102.4	175.6	66.5	53.9	2458	ohc	four	122	mpfi	3.31	3.54	8.7	92	4200	27	32	11248
-1	toyota	gas	std	four	hatchback	fwd	front	102.4	175.6	66.5	53.9	2414	ohc	four	122	mpfi	3.31	3.54	8.7	92	4200	27	32	9988
0	honda	gas	std	two	hatchback	fwd	front	96.5	167.5	65.2	53.3	2236	ohc	four	110	1bbl	3.15	3.58	9	86	5800	27	33	7895

0	toyota	gas	std	four	sedan	fwd	front	95.7	166.3	64.4	53	2094	ohc	four	98	2bbl	3.19	3.03	9	70	4800	38	47	7738
0	honda	gas	std	four	sedan	fwd	front	96.5	175.4	62.5	54.1	2372	ohc	four	110	1bbl	3.15	3.58	9	86	5800	27	33	10295
2	volkswagen	diesel	std	two	sedan	fwd	front	97.3	171.7	65.5	55.7	2261	ohc	four	97	idi	3.01	3.4	23	52	4800	37	46	7775



TEST DATA

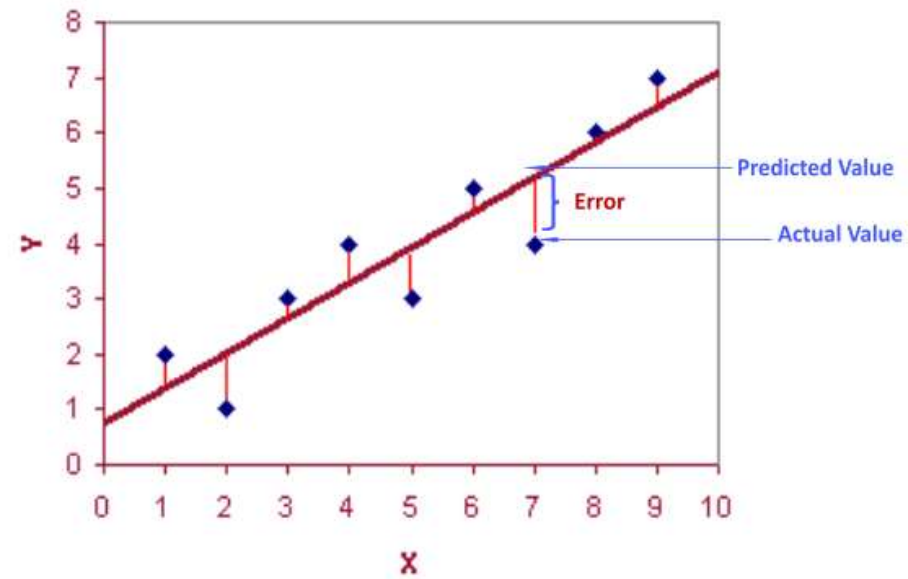


VOORSPELLING OP TEST DATA

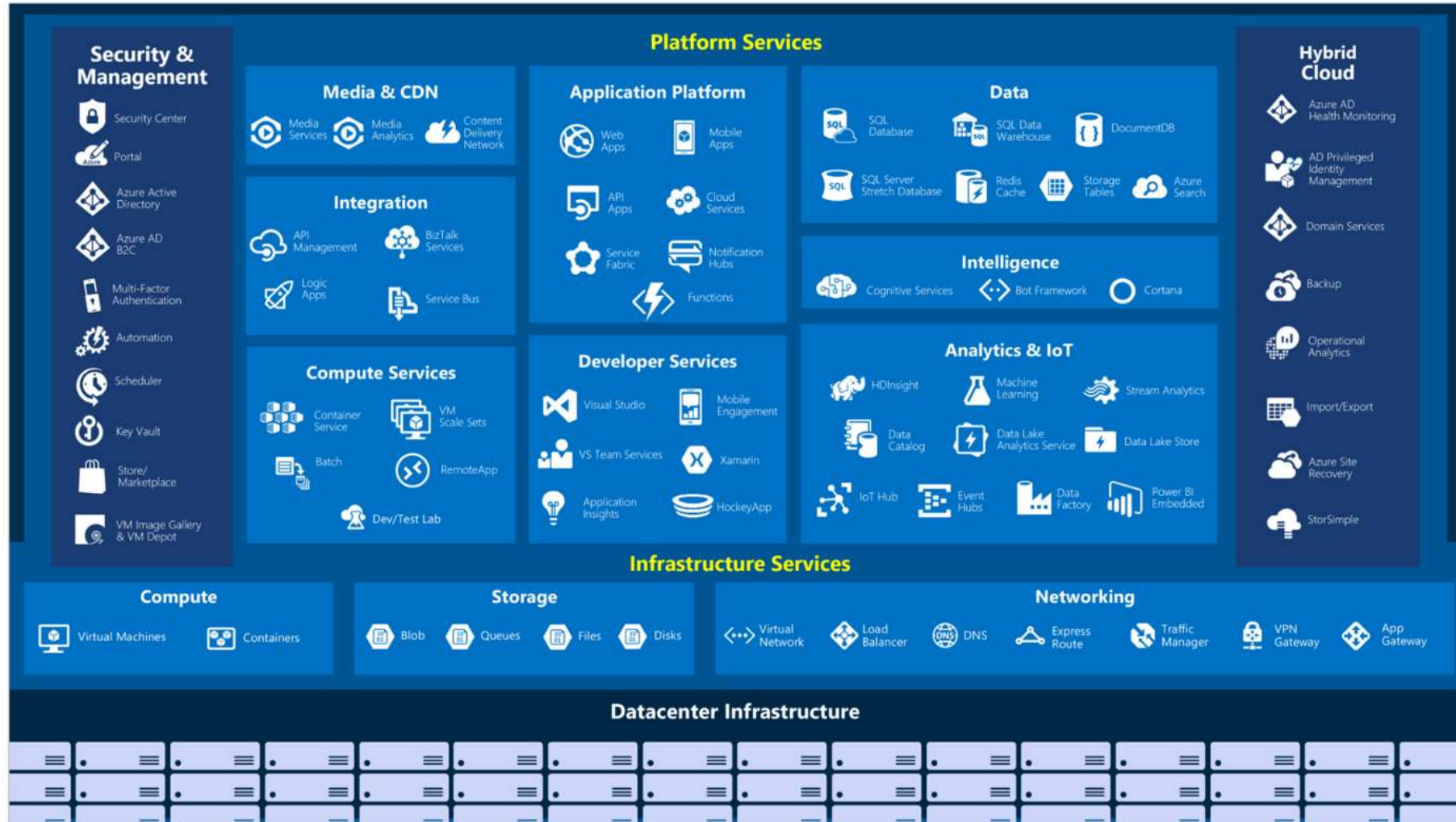
symboling	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	length	width	height	curb-weight	engine-type	num-of-cylinders	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price	Scored Labels
0	toyota	gas	std	four	sedan	fwd	front	95.7	166.3	64.4	53	2094	ohc	four	98	2bbl	3.19	3.03	9	70	4800	38	47	7738	7457.375
0	honda	gas	std	four	sedan	fwd	front	96.5	175.4	62.5	54.1	2372	ohc	four	110	1bbl	3.15	3.58	9	86	5800	27	33	10295	8918.75
2	volkswagen	diesel	std	two	sedan	fwd	front	97.3	171.7	65.5	55.7	2261	ohc	four	97	idi	3.01	3.4	23	52	4800	37	46	7775	9003.375

PERFORMANCE METRICS

PERFORMANCE METRICS | REGRESSION



AZURE ML

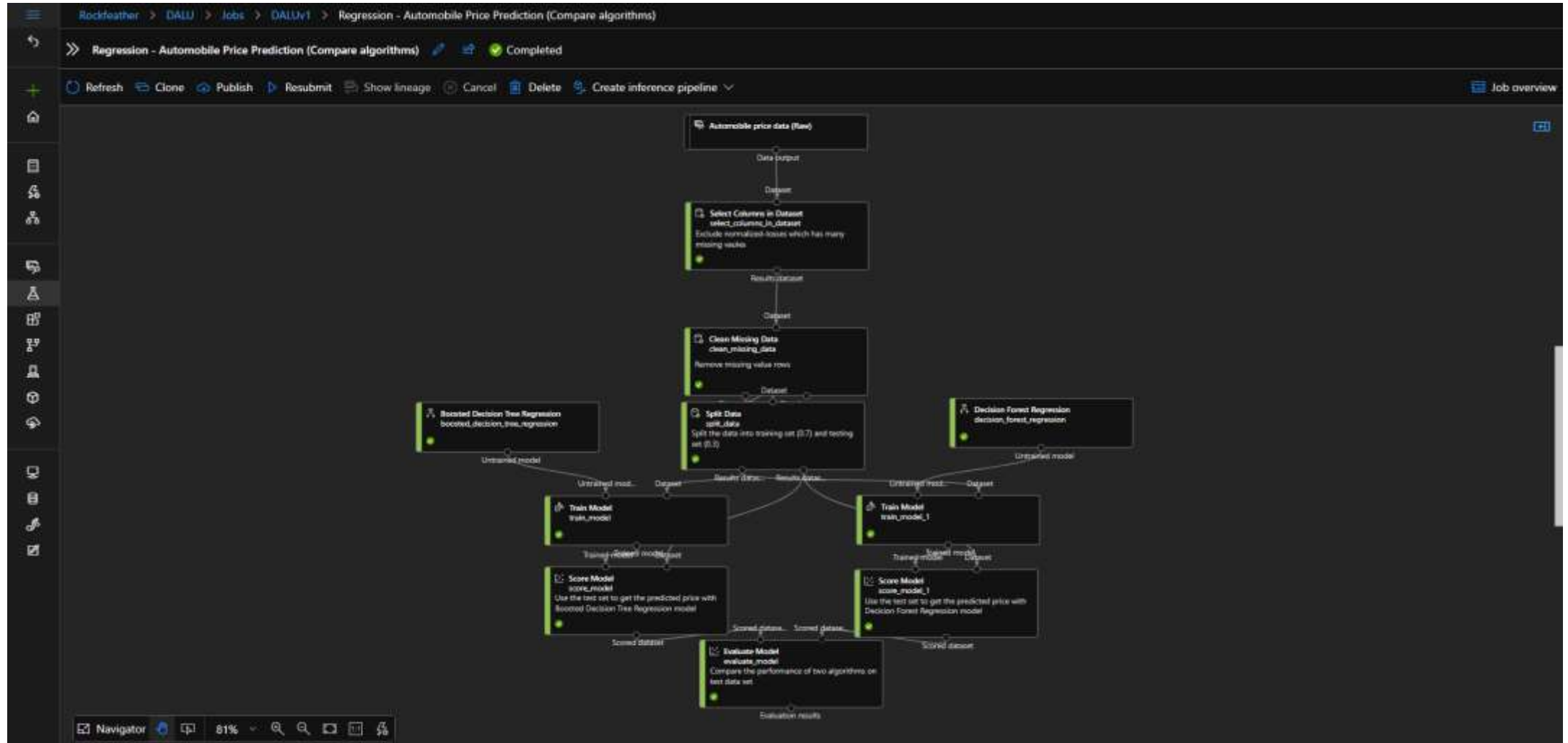


AZURE MACHINE LEARNING WORKSPACE

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes the Microsoft Azure Machine Learning logo and a search bar. The left sidebar contains a navigation menu with options: Home, Author (Notebooks, Automated ML, Designer), Assets (Data, Jobs, Components, Pipelines, Environments, Models, Endpoints), and Manage (Compute, Datastores, Linked Services, Data Labeling). The main content area is titled 'Rockfeather > DALU' and features a 'DALU' header with a pencil icon. Below this, there are four cards: 'Create new' with a plus icon, 'Notebooks' (Code with Python SDK and run sample experiments), 'Automated ML' (Automatically train and tune a model using a target metric), and 'Designer' (Drag-and-drop interface from prepping data to deploying models). Each card has a 'Start now' button. Below these cards is a 'Recent resources' section with tabs for Jobs, Compute, Models, and Data. The 'Jobs' tab is active, showing a table of recent jobs.

Display name	☆	Experiment	Status	Logs	Submitted time	Submitte...	Job type
keen_hair_pymynd8		automl-bi...	✓ Completed		Oct 10, 2022 10:...	Niels Jon...	Command
helpful_drop_mjs5rpmx		automl-bi...	✓ Completed		Oct 10, 2022 9:5...	Niels Jon...	Automated...
musings_atemoya_qhl0...		a0d6009a...	● Not started		Oct 7, 2022 1:14 ...	Niels Jon...	
Stap 7		DALU	✓ Completed		Oct 7, 2022 1:10 ...	Niels Jon...	Pipeline
Stap 7		DALU	✓ Completed		Oct 7, 2022 11:3...	Niels Jon...	Pipeline

AZURE MACHINE LEARNING DESIGNER



A person wearing a blue denim shirt is shown from the chest down to the waist. They are rolling up the left sleeve of the shirt. The background is a blurred, light-colored wall. A vertical red line is positioned on the left side of the image, to the left of the text.

DEMO | AZURE ML

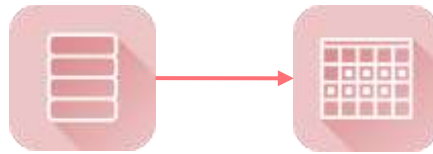
AUTO PRIJS CASUS

- Inladen data



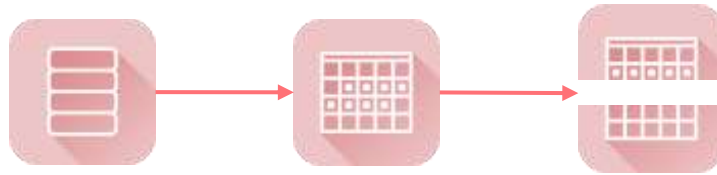
AUTO PRIJS CASUS

- Inladen data
- Opschonen van data



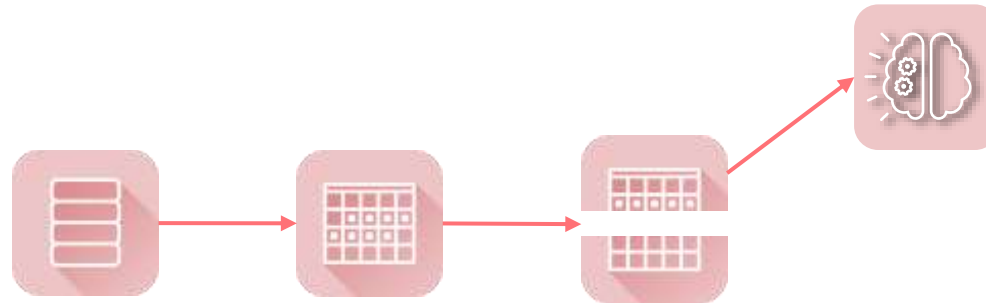
AUTO PRIJS CASUS

- Inladen data
- Opschonen van data
- Data splitten



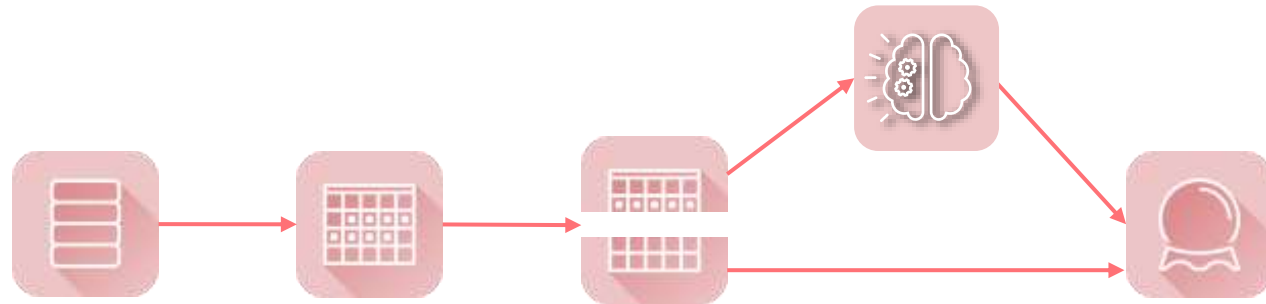
AUTO PRIJS CASUS

- Inladen data
- Opschonen van data
- Data splitten
- Machine learning modellen trainen



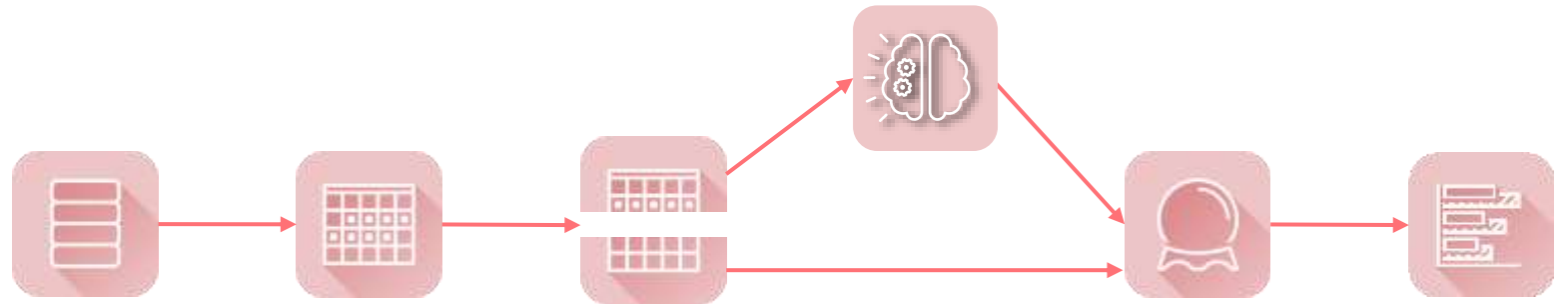
AUTO PRIJS CASUS

- Inladen data
- Opschonen van data
- Data splitten
- Machine learning modellen trainen
- Machine learning modellen testen



AUTO PRIJS CASUS

- Inladen data
- Opschonen van data
- Data splitten
- Machine learning modellen trainen
- Machine learning modellen testen
- Modellen evalueren



MACHINE LEARNING CANVAS



Value Propositions

What are we trying to do for the end-user(s) of the predictive system? What objectives are we serving?

- Onze verkopers helpen met het inschatten van de verkoopprijs van een auto
- Winst vergroten door beter inschatting van de waarde



Data Sources

Which raw data sources can we use (internal and external)?

- Openbare dataset
- Eigen historische verkoop data



Data Collecting

How do we get new data to learn from (inputs and outputs)?

- Openbare dataset downloaden
- Eigen data uit bron systeem halen



Data Engineering

Input representations extracted from raw data sources.

- Kolommen met veel missende data weggooien
- Rijen met veel missende data weggooien



Live Evaluation & Monitoring

Methods and metrics to evaluate the system after deployment, and to quantify value creation.

- Helpt van de verkopers laten werken met de voorspellingen en helpt zonder en na een half jaar kijken of de gemiddelde winst hoger was bij de groep die de voorspellingen gebruikte



ML Task

Input, output to predict, type of problem

- Prijs voorspellen
- Regressie probleem



Local Evaluation

Methods and metrics to evaluate the system before deployment.

- RMSE score om verschil tussen voorspelling en daadwerkelijke waarde te evalueren



Decisions

How are predictions used to make decisions that provide the proposed value to the end-user?

- Een goede voorspelling helpt de verkopers om de juiste prijs te bieden bij het kopen en verkopen van auto's

NEXT STEPS

1. Lees onze data science blogs op **rockfeather.com**
2. Schrijf je gratis in voor een Data Science Discovery Session, waar we samen het ML canvas invullen
3. Kijk de volgende data science webinars terug op Rockfeather.com:
 - Webinar: Predictive Maintenance
 - Webinar: Predict Churn with Artificial Intelligence



Niels Jonker

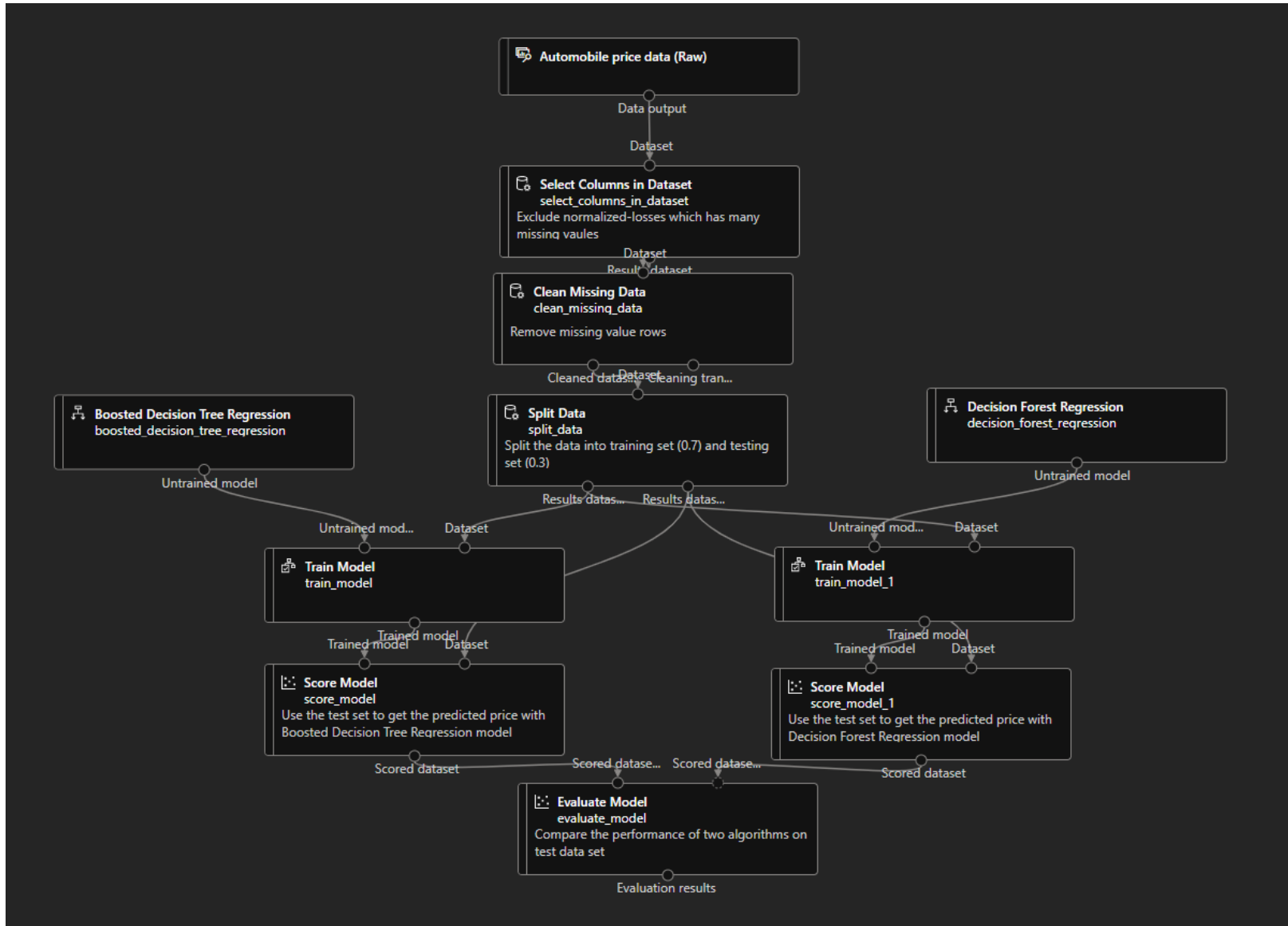
+31 6 18774517

niels.jonker@rockfeather.com

<https://www.linkedin.com/in/nielsjonathanjonker/>

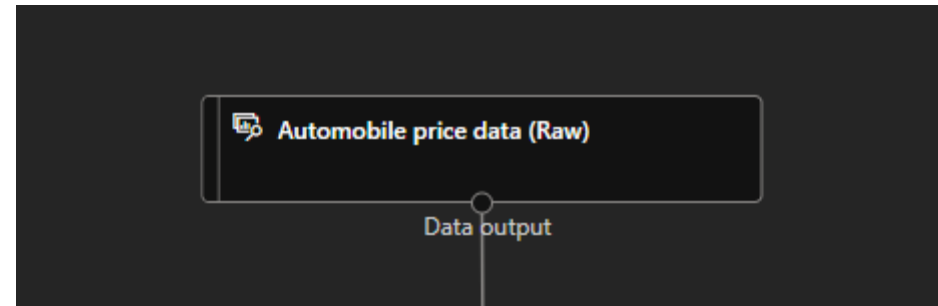


AUTO PRIJS CASUS



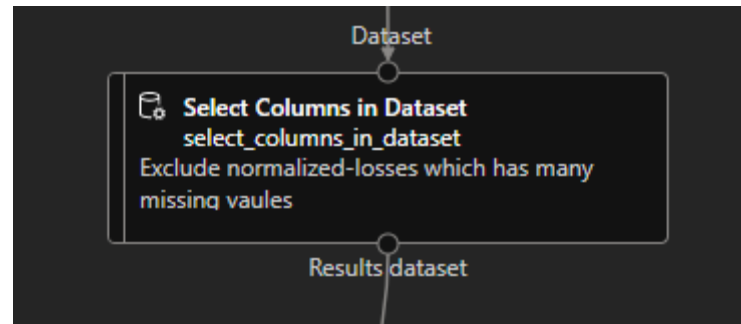
AUTO PRIJS CASUS

- Auto prijzen dataset (dataset van UCI Machine Learning Repository)



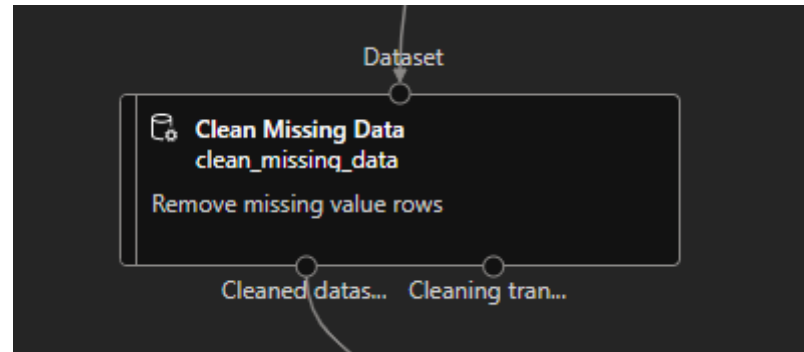
AUTO PRIJS CASUS

- Auto prijzen dataset (dataset van UCI Machine Learning Repository)
- Inladen data



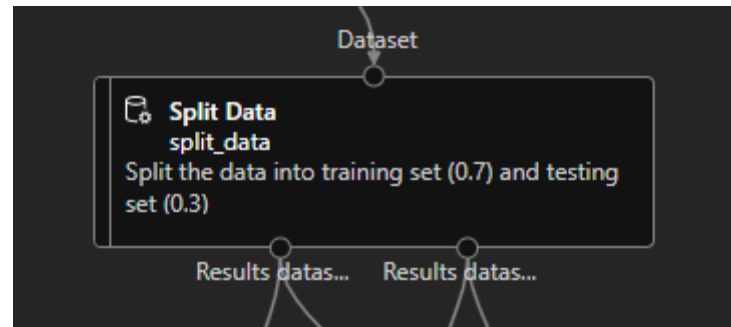
AUTO PRIJS CASUS

- Auto prijzen dataset (dataset van UCI Machine Learning Repository)
- Inladen data
- Opschonen van data



AUTO PRIJS CASUS

- Auto prijzen dataset (dataset van UCI Machine Learning Repository)
- Inladen data
- Opschonen van data
- Data splitten



AUTO PRIJS CASUS

- Auto prijzen dataset (dataset van UCI Machine Learning Repository)
- Inladen data
- Opschonen van data
- Data splitten
- Machine learning modellen trainen



AUTO PRIJS CASUS

- Auto prijzen dataset (dataset van UCI Machine Learning Repository)
- Inladen data
- Opschonen van data
- Data splitten
- Machine learning modellen trainen
- Machine learning modellen testen



AUTO PRIJS CASUS

- Auto prijzen dataset (dataset van UCI Machine Learning Repository)
- Inladen data
- Opschonen van data
- Data splitten
- Machine learning modellen trainen
- Machine learning modellen testen
- Modellen evalueren

